ROLE OF SEMANTIC WEB IN THE CHANGING CONTEXT OF ENTERPRISE COLLABORATION

by

Nitesh Khilwani

A thesis submitted in partial fulfilment of the
requirements for the degree of

Doctor of Philosophy

Loughborough University

October 2010

ABSTRACT

The concept of *virtual enterprise* is one of the most competitive industrial strategies for structuring and revitalising enterprises for the 21st century. In such collaboration, enterprises temporarily share competences and resources for a particular business goal and disband when that window of opportunity is closed. The paradigm of the virtual enterprise is a predominant area of research and technological development for today's progressive industries.

The developments in information technology and communication are the enabling factors for distributed companies to collaborate electronically and form virtual collaborations. However, the legacy systems running at present in the enterprises have not been developed to directly and independently connect with the heterogeneous systems of other enterprises. Moreover, the enterprise information resources are not structured with standard mechanisms, which can be readable and understandable to both humans and machines without any risk of misinterpretation.

Considering the present technological scenario and bearing market requirements in mind, this research aims for studying different types of virtual collaborations, identifying information sharing and knowledge management issues in such collaborations and subsequently proposing semantic frameworks to resolve those problems. The objective of this research is to propose semantic tools and techniques for providing explicit semantic description to the information resources shared and exchanged among heterogeneous enterprises in different stage of virtual collaboration.

In this research, the life cycle of virtual enterprise is divided into 4 stages, namely- Identification, Creation, Operation and Termination. Initially an ontology, ECOS (Enterprise Competence Organization Schema), is proposed in the identification stage to capture the skills and capabilities of the enterprises. This ontology is then used in the creation stage to identify relations among enterprises and select appropriate business opportunities and partners for them.

In the operation stage, an information retrieval tool, TEXT2RDF, is developed to extract information from the documents and convert them into a computer understandable resource using semantic web standards. This technique is also utilized in the termination stage to identify relationships between documents and reports available in the enterprise information repository.

Finally, a real life case study of a virtual enterprise is taken to illustrate the applicability and effectiveness of the semantic tools and techniques proposed for each stage of the virtual enterprise.

Keywords: Enterprise Collaboration, Virtual Enterprise, Ontology modelling, Semantic web, Information retrieval, Organizational Networks, Semantic annotation.

PREFACE

This thesis is submitted to the Loughborough University for the doctoral degree – Doctor of Philosophy (Ph.D.). The work has been carried out at the Knowledge Management for Manufacturing Group (KMM), Wolfson School of Mechanical and Manufacturing Engineering, under supervision of Dr Jenny Harding. This research was funded by Innovative Manufacturing and Construction Research Centre (IMCRC), Loughborough University.

Part of my PhD education in making this thesis has been taken place in I-Shou University (Taiwan), FZI (Germany) and Iowa University (USA). With the Overseas Research Scholarship provided by IMCRC, I visiting research groups in these universities to understand their expertise and incorporate them in my PhD research.

Acknowledgements

The thesis marks the end of a long journey where I first of all would like to express my sincerest gratitude to my supervisor Dr Jenny Harding for her thoughtful guidance and critical reviews to help me find a path when I got lost. It is an honour for me to pursue my research under her learned supervision and I will be obliged throughout my life for her kindness and support.

I would like to thank sincerely to Prof M. K. Tiwari, who helped me to make progress in the right direction when it was necessary. He definitively inspired me in many ways, thanks to his valuable expertise in the computer and security-related domains. Also, I would like to thank Prof Keith Popplewell for allowing me to work with him and be a part of SYENRGY project team.

I am lucky to have knowledgeable, friendly and helpful colleagues around during my work in KMM-group. I especially acknowledge Rahul Swarnkar, Alok Choudhary, Sri Krishna Kumar and Nagesh Shukla for the productive collaboration, inspiring and constructive discussions and the generous and sincere help in my research.

I also acknowledge the IMCRC at Loughborough University and the SYNERGY - FP7-ICT-2007 project members for their support during this research.

Finally, I would like to express my heartfelt appreciation and love to my parents and family for their support and understanding.

Nitesh Khilwani
October 05, 2010

CONTENT

LIST OF FIGURES

LIST OF TABLES

GLOSSARY OF TERMS

BOP	Bag of Words
BOW	Bag of Phrases
BPML	Business Process Modelling language
BusCO	Business Concept Ontology
CIM-OSA	Computer integrated manufacturing open system architecture
CM	Corporate Memory
CNO	Collaborative Network Organization
CPC	Central Product Classification
CPU	Central Processing Unit
CPV	Common Procurement Vocabulary
CRAI	Competence-Resource-Aspect-Individual
CRCTOL	Concept Relation Concept Tuple based Ontology learning
ECM	Enterprise Competence Model
ECOS	Enterprise Competence Organization Schema
FOAF	Friend-of-a-Friend
GLOBEMEN	Global Engineering and Manufacturing in Enterprise Networks
HTTP	HyperText Transfer Protocol
IMCRC	Innovative Manufacturing and Construction Research Centre
IPhrase	Important Phrases
ISCO	International Standard Classification of Occupations
ISIC	International Standard for Industrial Classification
IT	Information technology
KM	Knowledge Management
KMM	Knowledge Management for Manufacturing
LOD	Linking Open Data
LSA	Latent Semantic Analysis
MDC	Meta Data Coalition
NLP	Natural Language Processing
NUTS	Nomenclature of Territorial Units for Statistics

OIM	Open Information Model
OJEU	Official Journal of the European Union
OML	Ontology mark-up language
ONA	Organizational Network Analysis
OWL	Web Ontology Language
PCFG	Probabilistic Context Free Grammars
POS	Parts Of Speech
RDF	Resource Description Framework
RDF-S	Resource Description Framework Schema
SHOE	Simple HTML Ontology extension
SKOS	Simple Knowledge Organization Scheme
SME	Small and Medium sized Enterprise
SNA	Social Network Analysis
SOM	Self organized mapping
S-P	Subject-Predicate
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Sequential Query Language
SVD	Singular Value Decomposition
TOVE	TOronto Virtual Enterprise
UECML	Unified Enterprise Competence Modelling Language
URI	Unique Resource Identifier
vCard	Versit card
W3C	World Wide Web consortium
WWW	World Wide Web
XOL	Ontology exchange language

INTRODUCTION

Research is what I'm doing when I don't know what I'm doing.
Werner von Braun

1.1. Background

The current global market is characterized by dramatic and often unanticipated changes. Enterprises are operating in an environment where markets frequently change, new technologies continually emerge and competition is fierce at the global scale. The prime reason for this is the competition between large corporations and small and medium enterprises for the control of profitable business areas. Small enterprises are scrambling to establish their presence in the market, but are held back in collaborations or interactions with other companies by incompatibilities with their legacy systems. Reinvestment is difficult as they are often confronted by a lack of resource and finance that is uncommon for the larger companies and multi-national corporations. In this difficult and ever-changing milieu, the adoption of new information and knowledge management strategies is imperative if enterprises are to remain competitive in the market.

In order to compete with global giants, smaller enterprises may concentrate on their core competences and collaborate with companies that compliment their skills and core activities. Enterprise collaboration can be a beneficial strategy for companies to adopt, resulting in massive reductions of fixed cost, overheads and manpower, in some circumstances perhaps even enabling the survival of companies. However, it requires integration of applications and data sources so that they can easily share business processes and information system (Linthicum 1999). Basically, the concept of enterprise
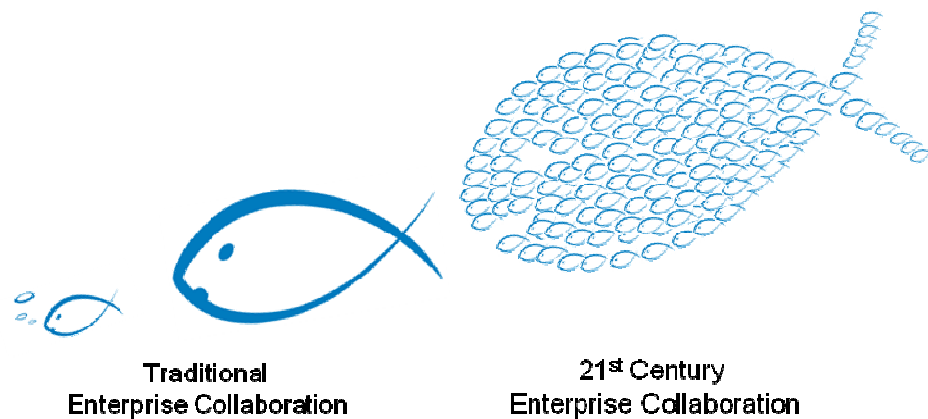
Figure 1-1: Two different models of Enterprise Collaboration

collaboration started in the early 90's, when dominant firms extended their boundaries and developed relationships with small companies, customers and services. In an extended enterprise model, the leading company focus on their core competences and outsource other business and technical activities to outside suppliers and other service providers. This model of extending an enterprise forced partner companies to compromise their independence and focus on business functions rather than on their fundamental competencies.

The negative aspects of the extended enterprise model lead companies to adopt the concept of a virtual collaboration. The concept of virtual collaboration is one of the most competitive approaches for enterprises wishing to enter global markets in collaboration whilst still maintaining their independence and autonomy. This 21st century enterprise collaboration is a temporary alliance of independent enterprises in which companies come together to share skills, core competencies and resources in order to better respond to business opportunities. In such virtual collaboration, enterprises temporarily share competencies and resources to respond to business opportunities in a more collaborative rather than competitive manner. A self explanatory picture representing the two different model of enterprise collaboration is shown in Figure 1-1.

1.2. Motivation

Virtual collaboration is not the end of the story, rather, it is the beginning of a new era with new challenges and pitfalls (Muata, et al. 2004). The IMS project, GLOBEMEN (Global Engineering and Manufacturing in Enterprise Networks), defined a virtual collaboration as a next generation enterprise with a network of globally distributed firms operating in an environment of abrupt and often unanticipated change (IMS 1996, YIT Corporation Jan 2000). The business partners are integrated using information and communication technology, such as online service, internet etc. Information technology can be considered as an essential strategy to store, publish, promote and share skills and abilities, and carry out collaboration efficiently. Advances in information technologies have strongly and consistently supported organizations to deliver *the right information to the right person at the right time*. This technology has weaved a pervasively networked world with millions of companies, billions of peoples and trillions of processes interconnected for handling the challenges of a network economy. Despite the increase in number of information entities (e.g. people, groups, organizations, etc) adopting such technological developments and joining the global market, little is known about companies with different organizational backgrounds, languages, customs, habits and locations.

In a virtual collaboration, information resources are generally distributed between the collaborating partners, rather than being centralized. Amalgamating such information for sharing and exploiting across multiple business functions and processes clearly has attractions since it reduces the potential for errors caused by data duplication and the resulting inconsistencies. But diverging goals, objectives and operational disconnects among collaborating firms make this information sharing and communication more complex. The next generation enterprise community needs sets of interconnected data and semantic models to communicate and exchange their knowledge, without the current risk of misinterpretation. The latest technological advancements are needed and explored in this research to share and effectively utilize a full range of data, information and knowledge.

1.3. Research Idea

Although, information sharing and maintenance within a co-operative environment has been considered for decades, it has always been an expensive and risky proposition (Lai 2007). It is only with the advent of web technology that information sharing has emerged as a viable technological solution. Web technology has enabled firms to make their information available electronically so that users and application programmers can access and share their resources and expertise. Despite growing interest and efforts, this technology is still primitive in its functionality. Today's web arranges the information syntactically, assuming it to be semantically homogenous which can commonly cause problems and misunderstandings (Casey, Pahl 2003). This downside means that most of the information must be interpreted by humans before use, rather than being readily available for use by machines. The availability of information that requires human interpretation does not necessarily solve the complex problem of getting *the right information to the right person at the right time*. At best, the need for human interpretation slows the information transfer down, and in worst case humans are prone to errors in understanding and interpretation.

Information sharing and communication among multidisciplinary enterprises require the treatment of structure as well as semantics of the data stored in enterprise data and records repositories (Casey, Pahl 2003, Xu, Xiao 2007, Lastra 2006). Collaborating partners often use different terminologies to describe the same meaning or alternatively the same terminology may be associated with different meanings. Such lack of clear semantic description of the meaning of the data's contents in a particular domain has hampered the interoperability and sharability of information between collaborating firms. In order to handle this interoperability challenge of future organizations, it is necessary to make information explicit and machine interpretable (Vernadat 2002, Pépiot, et al. 2007). To overcome this problem, researchers are focusing on semantic web concepts and tools that enable computers to automatically process and understand the information (Zhou, Dieng-Kuntz 2004).

The Semantic Web is a vision for extending current web technology in which information is annexed with a well defined meaning that enhances the interoperability of computers and people (Casey, Pahl 2003). The primary benefit of this new vision of Tim Berners-Lee et al. is to represent the web resources in formalisms that both machines and humans can understand, thereby unleashing the potential for software agents to perform tasks on behalf of humans (Berners-Lee, et al. 2001) The concept of the semantic web is, without any doubt, gaining attention in both industry and academia. The growing access to heterogeneous and independent data repositories in enterprise collaboration is attracting a lot of research communities to identify methods for representing data in a way that it can be shared and processed automatically on the semantic web (Terziyan, Kononenko 2003).

Keeping the present market and technological scenario in mind, this research aims to determine a platform that enables companies to achieve their potential through semantic web technology and assimilate distributed information resources. The idea is to provide explicit semantic descriptions to the enterprise information resources and encode them in an unambiguous and machine understandable form. In this research, two types of information resources, i.e. Business Profile and Text documents are considered and converted into semantic web resource. The resulting semantically enabled information is then be exploited and disseminated through the proposed platform to project teams and organizations collaborating in a specific problem domain.

C h a p t e r   2

ENTERPRISE COLLABORATION: CONCEPT AND TREND

The secret is to gang up on the problem, rather than each other.
Thomas Stallkamp

ABSTRACT

The current trend in the market is to develop temporary alliances of independent enterprises, in which companies come together to share skills, core competencies and resources for short or long term collaborations. The concept of a virtual enterprise is one of the most competitive approaches for enterprises wishing to enter the global market for collaboration whilst still maintaining their independence and autonomy. This chapter studies different styles of virtual collaboration and its life cycle.

CHAPTER OUTLINE

2.1.   What is Enterprise?
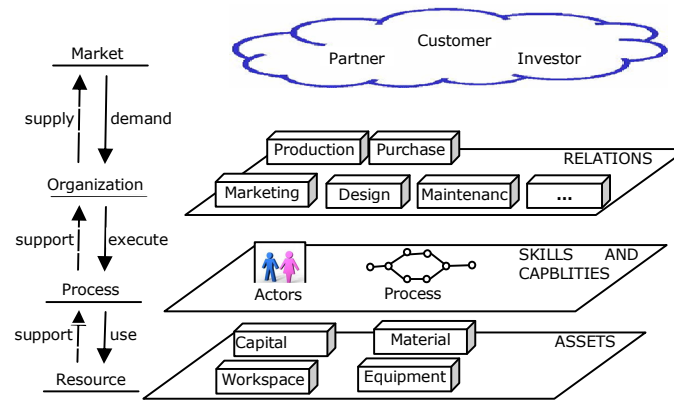
Figure 2-1: Hierarchal Structure of an Enterprise

An enterprise is a complex and dynamic management system comprising of all facets of business, including purchasing, marketing, design, production, maintenance and distribution etc (Yu, et al. 2000). It can be considered as a myriad of concurrent processes executed by a huge set of technological and/or human entities. The hierarchical structure of an enterprise is shown in Figure 2-1. The processes, resources and strategies of an enterprise are managed as per their business missions and are subject to the internal and external constraints due to the market. Enterprise activities are arranged in a manner that supports the delivery of superior quality products and services, rapid market response and at the same time keeps the production cost as low as possible (Fox 1993).

The concept of enterprise began in the 20th century, when major companies such as Ford, Toyota, HP, Microsoft, etc came to the market with their specific products/services. Initially enterprises operated in an environment with clear boundaries, limited relationships with others and a strong internal focus on efficiency and quality. However, in the 1980s, companies realized the meaning and significance of collaboration and started sharing business and technical activities with outside suppliers and service providers (Sun Microsystems 2004). To cope with the anticipated complexities, firms started looking beyond their self centred passive environment and developed collaborations with organizations that complement their skills and core activities. This

step was an impressive and innovative move for companies to make with the aim of reducing the burden of costs associated with the development of non-core activities. In such environments, organizational barriers broke down, partnerships with suppliers, clients and even competitors were common and had to be managed with care, and efficiency and quality was considered beyond enterprise boundaries. The trend among enterprises was to form collaborative networks to succeed and achieve business goals by (Linthicum 1999)(Cross, et al. 2007):

•   building efficient system integration and business coordination,

•   utilizing collaborative work for business processes and organizational activities,

•   enhancing flexibility throughout the enterprise,

•   developing interoperability of techniques, systems and people to face market variability in a cost effective manner.

2.2. Enterprise Collaboration

With the advent of technological and economical globalization, numerous formal and informal networks started emerging among organizations and individuals (Cross, et al. 2007). The 'Automotive industry' is one of the best examples showing this transition. In order to compete with the global giants, the automotive sector concentrated on core competencies and teamed up with organizations that compliment their skills and core activities. Companies adopted this new policy by sharing designs, development and platforms and thus reduced the burden of costs associated with the development of non-core activities (Gialelis, et al. 2005). This revolution started in the recession of the early 90's, with a massive move towards outsourcing of non-core activities such as information technology, public relations, human resources etc. This was a successful strategy in the automotive sector that resulted in massive reductions of fixed cost, overheads and manpower as well as improved survival of companies.

The concept of enterprise collaboration brought a new insight in terms of business model, with enterprises shifting from self centred closed markets to a global open environment. Extended enterprise, virtual enterprise, supply chain, dynamic alliance and

industrial alliance are all distinguishable forms of competitive strategy (through collaborative networks) that have been explored for enterprise collaboration (Browne, Zhang 1999). The two main approaches that are often used to describe these networked environments and which are the main focus of this research are: extended enterprise and virtual enterprise. Both these terms are used in the context of enterprise collaboration for facilitating cooperation and interoperability among customers, suppliers, service providers, along with government agencies, academic organizations etc. However, the concept of an extended enterprise focuses on long term enterprise relationships, whereas the virtual enterprise creates collaborative alliances for relatively short periods of time. A comparative study on these two collaborative models is as follows:

- *Extended Enterprise*: An extended enterprise is a kind of collaboration, in which an enterprise extends its boundaries to encompass its suppliers, customers and other business partners. According to (Ratchev, et al. 2000),

  > The term extended enterprise represents the concept that a company is made up not just of its employees, its board members, and executives, but also its business partners, its suppliers, and even its customers. The extended enterprise can only be successful if all of the component groups and individuals have the information they need in order to do business effectively.

  In such collaborations an enterprise focuses on its core competences and outsources all other business and technical activities to outside suppliers and other service providers (Harding, Popplewell 1999). This concept is generally implemented by dominant enterprises to develop long term relationships with key customers and service providers, and treat them as important business partners.

- *Virtual enterprise*: A virtual enterprise is a temporary network of independent companies collaborating for a particular business goal. It is a collection of legally

Table 2-1: Comparison between Extended Enterprise and Virtual Enterprise (Browne, Zhang 1999)

S. No.	Criteria	Extended enterprise	Virtual enterprise
1	Strategic issues	Long term objective	Relatively short term
2	Partnership purpose	Business co-operation	Temporary collaboration for project or product
3	Organization stability	Stable collaboration of firms	Dynamic organization of companies, each with their core competencies
4	Relationships	Trust and mutual dependence	Temporary and dynamic
5	Coordinator	Generally the manufacturer or any other strong lead organization	Managed in accordance with an agreed collaboration contract or agreement
6	Information technology (IT)	Facilitated and enabled by IT	Enabler for the cooperation
7	Organization type	Dependent on coordinator	Generally, independent bodies cooperating for a particular mission

independent enterprises, institutions or single persons who come together to exploit fast changing opportunities. A virtual enterprise is defined as (Luczak 2005):

> A temporary alliance of independent enterprises that come together to share skills, core competencies and resources in order to better respond to business opportunities and whose cooperation is supported by computer networks.

In a virtual enterprise, companies are temporarily linked in a network to share skills, costs and market. The cooperating nodes in such an environment are geographically distributed and coordinating through electronic communications. The business partners are integrated using information and communication technology, such as online service, internet etc. They have no central office to control the working of firms and are a flat organization with no hierarchy and no vertical integration. Each partner is the owner of his enterprise and directly contributes to its success and failure. A comparative study of extended and virtual enterprises is shown in Table 2-1 (Browne, Zhang 1999, Zhang 1998).

2.3. Virtual enterprise: 21ˢᵗ Century Enterprise Collaboration

*The central difference between old and new economies: the old industrial
economy was driven by economies of scale; the new information economy is
driven by the economics of network.*

Carl Shapiro, Hal R. Varian

The shift from industrial economy to information economy in the current global market has brought an enormous increase in competitiveness, leading to the adoption of new organizational models (Shapiro, Varian 1998). The virtual enterprise model is one of the best strategic moves for enterprises seeking for global reach without compromising their agility and independence (Canavesio, Martinez 2007). The term *virtual enterprise* was coined in the 1990s, when communication technology and management theory compounded to focus business strategy on core competences and outsourcing of all non-core activities. It differs from existing inter-organizational models by the degree of shared accountability and responsibility of the participants and the structure by which companies contribute their competencies.

According to Davidow and Malone (Davidow, Malone 1992), the virtual enterprise is the industrial strategy for structuring and revitalising companies for the 21ˢᵗ century. Forbairt defined virtual enterprise as a response to the speed and globalization of the current economy (Forbairt 1996). It is set up with an objective to fill a window of opportunity and dissolve when that window is closed. Zimmerman explained the notion of virtual enterprise as a computer memory architecture where an individual enterprise with main memory borrows extra memory from other companies when needed (Zimmermann 1997).

In principle, a virtual enterprise is a temporary network of independent companies, coming together to exploit the fast changing opportunities of the global age. Burn et al. defined three fundamental attributes of a virtual enterprise, i.e. (Burn, et al. 1999):

1. *Culture*: This attribute is a perception of the virtual enterprise that is defined by the common shared values and beliefs of the members of a collaborative alliance. Tushman and O'Reilly suggested that enterprises with a culture of accepting new

technologies and highly decentralized and change oriented outlook are more likely to embrace virtuality and proactively seek further opportunities both within and without the collaboration (Tushman, O'Reilly 2006).

2. *Network*: This is defined as groups within an enterprise and among enterprises giving paramount importance to the development and maintenance of communicative relationships. However, the ability to establish multiple alliances along with retaining individual identity creates a constant tension between autonomy and interdependence, competition and cooperation. The virtual networks are often seen as value added partnerships of competitors, collaborators and/or complementary providers of goods and services joined to achieve competitive advantage over an individual enterprise.

3. *Market*: Markets are traditionally associated with pricing mechanisms, but in a virtual enterprise, this refers to an electronic market where there is a prerequisite is to enhance the ability of any buyer or seller to interconnect with a network and offer/accept goods and services. The essential criteria for a successful market includes products with low asset specificity and ease of description and a consumer market willing to buy without recourse to visiting retail stores.

   Skyrme explained the paradigm of virtual enterprise as a virtual world with virtual products/services, virtual work places and virtual organizations that can bring the following benefits (Skyrme 1999):

- *Virtual product/service*: produce, deliver and sell through electronic networks and thus create opportunities for accessing a wide range of specialized resources and reaching distant markets easily.
- *Virtual workplace*: work location independent to reduce costs of various activities
- *Virtual organization*: carry out work in teams and beyond company boundaries to gain flexibility in work.
- *Virtual world*: develop a broad social environment for the establishment of the virtual enterprise.

Despite having the above advantages, there are various drawbacks of a virtual enterprise. The concept of virtual enterprise assumes that an enterprise investing as little in infrastructure as possible is more responsive to a change in the marketplace and is therefore more likely to attain global competitive advantage, but this ignores the very real fact that large organizations can deliver sustained innovation to the market over the longer term (Menagh 1995). Similarly, Kuttner pointed out other difficulties that are (Kuttner 1993):

- The firm is no longer a physical entity with a stable mission, but a dynamic entity involved in a temporary alliance via networks, phones and internet etc.
- The workers become freelancers, whose fate in the marketplace is determined not by loyalty and diligence, but by their knowledge and skills.
- Working in a virtual environment involves contingency risk and security risk.

2.4. Different Styles of Virtual Enterprise

The concept of virtual enterprise has changed the self centred passive environment of an enterprise into a virtual collaborative environment of organizational networks linked for achieving business goals in a more productive and efficient manner. Camarinha-Matos and Afsarmanesh defined such a collaborative network as a set of nodes and edges linked by certain types of relation and flow (Camarinha-Matos, Afsarmanesh 2005). Each node in a network corresponds to an enterprise operating as a supplier, customer, producer, distributor and/or other specialized service provider. An enterprise interacts with one or several nodes by defining certain relations based on common goals, types of partnership and governance rules, for the flow of resources, e.g. information, product, control, etc. The enterprise nodes are networked electronically utilizing information and communication technologies to enhance organizational activities by reducing the need for physical and formalized structures.

In literature, different forms of collaborative networks have been classified on the basis of the types of nodes, linkages, relations or flows (Camarinha-Matos, Afsarmanesh 2005, Rajsiri,, et al. 2007, ECOLEAD 2008). In an early attempt at virtual enterprise

classification, Camarinha-Mahtos et al. (Camarinha-Matos, et al. 1998) defined the following characteristics for classifying different perspectives of virtual enterprise:

- *Duration*: In terms of duration, there are two types of collaborative networks i.e. short term alliances made for single business opportunities that are dissolved at the end of a project, and long terms alliances that last for an indefinite number of business processes or for a specified time span.

- *Topology/Geometry*: The topology of a network is another way of characterizing a virtual enterprise with dynamic and fixed structure of networks. Dynamic networks are the most typical case of a virtual enterprise where enterprises can dynamically join or leave the alliance according to the phases of the business processes or other market factors. But in many sectors there are networks (e.g. supply chain) with little variation in terms of suppliers or clients.

  Another facet related to the *geometry* is the possibility of an enterprise participating simultaneously in various networks or being committed to a single alliance (exclusivity).

- *Purpose*: This factor classifies a virtual enterprise based on the motivation of an enterprise to join the network. An enterprise can join a network 1) to extend its boundaries keeping control over its vital suppliers or to complement its core competencies in order to be able to share some market opportunities, 2) to be involved in a consistent supply chain from raw materials to end customers or to bid for a single opportunity in the market, 3) to increase its geographical presence or to improve the quality and responsiveness to its market opportunities.

- *Coordination*: In terms of network coordination various models can be found in literature. Camarinha-Mahtos et al. suggested three basic structures for classifying networks i.e. star like structure, democratic alliance and federation (Camarinha-Matos, et al. 1998). Similarly, Katzy et al. proposed three different structures namely, chain, star and peer-to-peer structure (Katzy, et al. 2005). Tagg distinguished ten different styles of virtual enterprise: 1) formal joint venture, 2) informal supply chain with no dominant player, 3) supply chain with dominant player, 4) end client with

preferred suppliers, 5) agency with freelance operator, 6) organization with home-based outworkers, 7) marketing cooperative, 8) cartel, 9) professional association, 10) voluntary working group or committee (Tagg 2001). The first six styles can loosely be generalized as supply chain relationships with different roles for dominant partners and end clients. The next four types are primarily the networks where enterprises perform activities with similar interests. Burn et al. suggested six different models of collaborative networks operating along a continuum and within a dynamic framework of change (Burn, et al. 1999).

The six different virtual alliance models proposed by Burn et al. are explained in Table 2-2 (Burn, et al. 1999).

2.5.   Virtual Enterprise: Life Cycle

The development of a virtual enterprise is based on a perception that together partners can achieve some specific business goals that they could not undertake individually.   The primary challenge in creating and operating a virtual enterprise is the selection and implementation of appropriate business partnerships and processes. Once the collaboration is achieved its efficiency is determined by the speed and accuracy with which information can be managed and exchanged among the business partners. Finally, when the market opportunity is satisfied, the network is disbanded, and knowledge and experience is retained for future use.

The above process and challenges illustrate that the virtual enterprise, similarly to life systems, follow a life cycle of creation, development and disappearance. Kanet and Faisst designed a five phase life cycle of virtual enterprise, with identification, formation, design, operation and dissolution phase (Kanet et al, 1999). This life cycle was designed to clarify the role of broker (also known as coordinator or mediator) in a virtual enterprise. In this life cycle, during the first two phases the broker finds appropriate business opportunities and creates virtual collaborations respectively.

Table 2-2: Different Styles of Virtual Enterprises (Burn, et al. 1999)

Style of VE	Definition	Example	Figure
Virtual Face	This model can be actualized as e-shop, e-auction or e-mail. It uses ICT in existing non-virtual enterprises to create additional values e.g. enabling customers to carry out same transactions over the internet which they have to do using telephone or fax.	Flower selling, Ticket booking, Electronic procurement, etc	
Co-alliance	A shared partnership with each partner brining approximately equal amounts of commitment to the network.	Oil exploration or production consortium, Bank's ATM consortium, Research consortium, etc.	
Star alliance	Coordinated network of interconnected members lead by a dominant enterprise. The central partner holds the decision making power and manages the entire network. This type of network corresponds to an extended enterprise in which each partner provides key functionalities.	Generally large scale enterprises, e.g. Construction industry, Large automobile manufacturing industry.	
Value alliance	It is based on supply chain model where enterprises bring together a range of products, services and facilities in one package. Here, partners come together on a project by project basis but coordination is provided by the general contractor.	Trade associations, Production and distribution chains, Research collaboration.	
Market alliance	Such alliance primarily operates in electronic market where partner brings range of products and services, each of which may be provided by individual enterprises.	Amazon.com, Ebay, Confuse.com, etc	
Virtual broker	It is purpose built virtual enterprise with highest level of flexibility. It is created to fill a window of opportunity and dissolve when that window is closed.	Seasonal market, Freelancing, etc.	
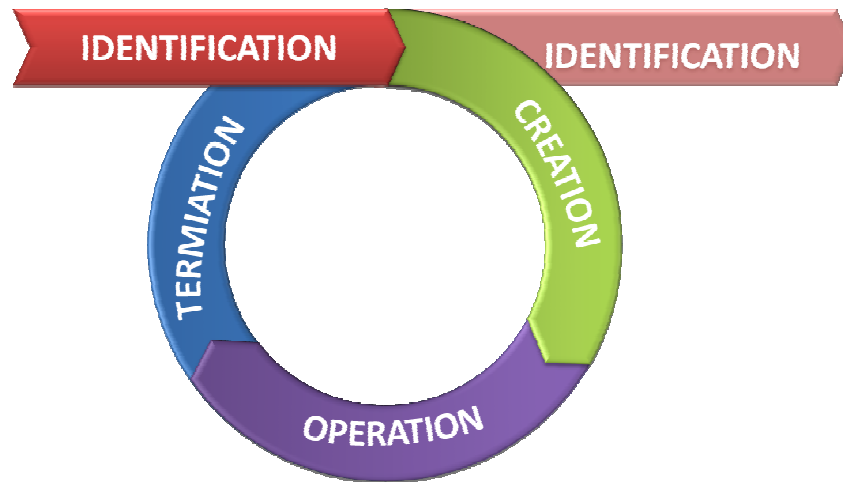
Figure 2-2: Life Cycle of Virtual Enterprise

During the design phase, the broker specifies the contractual agreements between virtual enterprise members. In the operation phase, broker acts as project moderator and finally during the termination phase the broker disbands the network.

However, in this research the focus is mainly on information sharing and knowledge management problems in a virtual enterprise. Therefore, the design phase in the above life cycle dealing with legal frameworks is not within the scope of this research. Hence, the life cycle of a virtual enterprise considered in this research, includes four stages, namely 1) Identification, 2) Creation, 3) Operation and 4) Termination, as shown in Figure 2-2. The problems focused in this life cycle are:

1. *Identification*: This is the stage where prospective companies are in pursuit of emerging business opportunities. Enterprises search for projects that can be handled individually or in collaboration with other companies. In order to develop a collaborative environment, a relatively stable network of potentially collaborating companies needs to be established to provide a basis for social relationships and trust among companies. The companies willing to develop such a network must publish information about their competences (knowledge that the enterprise is willing to make public) about what they are able to offer in a collaboration.

2. *Creation*: In this stage, enterprises develop a relatively stable network of partners for the collaboration. Such a network gives a basis for social relationships and trust among companies. In order to support collaboration, the new business opportunities must be identified, awareness should be created about the opportunities among the potential collaborating partners, suitable partners must be selected, and collaboration agreements and contracts must be finalized (Huang, Diao 2007).

3. *Operation*: The main objective during the operation phase is to deliver items on time as promised or as agreed with the customer and other collaborative partners. In a virtual collaboration, knowledge is distributed between the potentially collaborating partners, rather than being centralized. Distributed knowledge means that no one is capable of achieving the highest level goal(s) on his own; goal achievement becomes a team activity, where different enterprises share information for the achievement of goal(s). The partners within the virtual enterprise need to gain and share knowledge about the product or service they are trying to deliver to the customer. Although, working with teams from different backgrounds, sectors and working cultures may involve conflicts between the partners, which may lead to the conflicting decisions and unwanted results. These problems should be resolved through successful negotiations to make the project a success and maintain the business reputations of the associated partners.

4. *Termination*: Finally, at this stage the collaborative project is finished and the network is disbanded. In this phase, it is necessary to gather experience from project members so that it can be utilized in future projects. Although, the shared knowledge is unlikely to be needed in its shared format after termination, anything of future use should be retained by individual companies (Dieng, et al. 1999). Generally, the process and agreements about knowledge sharing and dissemination is decided during creation stage in collaboration contracts and agreements.

The key goals and processes involved at each stage of a virtual enterprise for different styles of collaborations are provided in Table 2-3. As discussed in Section 2.4. (Coordination), different styles of coordination models are developed in the literature. In

this research, the key research findings from the different models developed by Camarinha-Matos, et al., Katzy, et al., Tagg and Burn, et al. have studied and collected together in Table 2-3 (Camarinha-Matos, et al. 1998, Katzy, et al. 2005, Tagg 2001 and Burn, et al. 1999).

2.6. Virtual Enterprise: Information Sharing and Knowledge Management

The emerging trend of virtual enterprise collaboration imposes an increasing demand for autonomous collaboration, reconfigurability and heterogeneous system integration. The need is to effectively manage the present and shape the future, providing a consistent route towards continuous improvement. Information flow is a major requirement for effective and efficient working in a collaborative environment. From Table 2-3, it is evident that information sharing and knowledge management is the basic requirement for each stage of virtual enterprise. The availability of information, ability to exchange it seamlessly and process it quickly is a major requisite for making decisions in a collaborative environment. However, the growing complexity of information resources and the increasing amount of knowledge and information required by a wide variety of users has made it increasingly difficult to share and exchange knowledge efficiently and effectively. The difficulties in information sharing and collaboration caused by diversity in enterprise heterogeneity are as follows (Linthicum 1999, Vernadat 2002):

- Differences in the workflow cause possible difficulties in coordination of the knowledge transfer
- Differences in terms create misunderstandings and lead to a different interpretation in the knowledge exchange process
- Differences in organizational cultures cause difficulties in accessing the complete knowledge of the firm.
- Difficulties in structuring and classifying the companies knowledge causes misunderstanding in sharing of information.

Table 2-3: Key Goals and Processes for Different Styles of Virtual Enterprise

Style of VE	Goals	Processes			
		Identification	Creation	Operation	Termination
Virtual Face	Achieve stated aims while minimizing the effort needed by company and customer.	Identify appropriate parent enterprise and the virtual face	Decide terms and conditions for business; Discussion on publishing and securing information	Access the knowledge base of parent enterprise and fulfil customer needs.	End the access to information; Update customer database
Co-alliance	Create banner to do business; Share costs, risks and complementary skills.	Identify broad areas of opportunities and suitable partner	Decide tasks for individual partners and knowledge management infrastructure.	Design and execute workflow pattern and share information when requested	Learn lessons from project and update knowledge base
Star alliance	Dominant- Concentrate on core competences; Member- Reliable business	Identification by dominant player for potential partners; negotiation including bidding ; identity check	Dominant enterprise assigns tasks to partner enterprises.	Dominant- Evaluate people, project and task; Member- Update dominant player	Disband the members; Learn lessons from the project.
Value alliance	Trusted partners and secure business; Carry out projects that involve both core and non-core activities.	Identify potential enterprise; Create collaboration pool with identity check and/or recommendation	Identify new business opportunities; Choose project head and partners for project; Discuss task among partners with general agreement or bidding ;	Provide product/service and share information with enterprise concern; Include new partners as per the project needs.	Prepare project review documents; Learn lessons from the project; Look for new projects;
Market alliance	Reliable supplier/customer business without formal ties.	Form initial group and look for other business partners.	Discussion on publishing and securing the information; Negotiate terms and condition for business	Share information with related enterprise; Include new partners as per the market needs.	End the access to information; Update supplier and customer database;
Virtual broker	Study market needs and new market entrants; Maintain fee levels.	Study the market needs and identify suitable enterprise to fulfil the need	Discuss the lifespan of project and supply of product/service	Evaluate assigned task and project.	Disband the member and look for new market need.

Various tools and techniques have been proposed in literature to solve these difficulties and reduce diversity. For example, *expert systems* are proposed for decision making, *standards* are used for information sharing, *web technologies* are available for data publishing and access, *ontologies* are used for knowledge management, *semantic web i*s used for understanding and interpreting the terms, *web services* are for accessing the knowledge of the network (Linthicum 1999, Lin, Harding 2007). However, the focus of this research is mainly on information sharing and knowledge management.

For effective information exchange in distributed collaborative environments, it is necessary to maintain compatibility and preserve syntactic and semantic content during the exchange of information. Initially, offline technologies such as client server architectures and local area networks were used to enable communication from one specific application to another. However, an increased number of information entities and matured use of information technology in the enterprise community has made it difficult to provide separate programs for linking heterogeneous enterprise applications. Thus, offline knowledge sharing was replaced by real time technology i.e. Web technology and Internet. Web technologies have been widely employed in organizations to underpin pervasively networked work that interconnects entities in various ways.

Web technology and the internet have initiated and made possible the integration of diverse systems and applications both inside the enterprise as well as between different enterprises. Subsequently, web services were introduced and designed to improve the interoperability across companies and wide area network. Web services are the set of functionalities used to enhance the potential of web technology by providing ways for automated program communication and the discovery of services (Tidwell June, 2000). This ubiquitous technology has been widely employed for building a robust, sharable and extensible collaborative architecture over the internet. Moreover, it has provided new opportunities for firms to facilitate the sharing of resources and therefore allows them to enjoy the benefits of integration (Cerami 2002).

Traditionally, these services are made available to cope with the differences in structure and syntax of heterogeneous databases and web programs. This technology generally provides no information about the semantics of the described language. It is based on an individual programmer or interpreter understanding the semantics of the available description and guiding the computers. The abilities of web technology and services are enhanced by creating semantic mark-ups that make them machine understandable and user apparent. These mark-ups are built using ontologies that represent particular domains through their concepts and the semantic relations between them (Terziyan, Kononenko 2003). The key enablers associated with the enhancement of traditional web technology are as follows:

1. Semantic Web: This is the term coined by World Wide Web (WWW) and is defined as an extension to the traditional web technology (Casey, Pahl 2003). It describes methods and technologies for the conceptual structuring of the semantics of data in a machine readable way that enhances the understandability of disparate information (Guha, McCool 2003).

2. Ontology: An ontology is considered as the backbone of the semantic web and the keystone for the web's automated tasks such as searching, merging, sharing, maintaining, customizing, and monitoring. It is used for defining the semantics present in the concerned domain and for sharing them among people, databases and applications (Lammari, Métais 2004). It is used to provide machine processable semantics for data and information sources that can be used by an agent (software or human), application or other information resources.

3. Information Retrieval: The vision of semantic web is to associate machine processable metadata to describe the semantics of information entities and facilitate searching, querying and reasoning of knowledge for human users. However, it is not practical to associate an ontology with all the information resources, which exists due to the gigantic amount of them. To promote the web to the semantic web, automatic

and semi-automatic information retrieval tools are used for annotating information resources with metadata and tags (Lai 2007).

As parts of an effort to improve cooperation, communication and collaboration, enterprises are adopting semantic web technology to share and exchange unified and unambiguous information resources and interoperate at the data level. In this research, these technologies are exploited to facilitate enterprise communication and collaboration.

2.7. Research Gap

The importance and necessity of semantic web technology has been recognized and widely accepted by industry and academic research. However, the two worlds have proposed solutions that progress along different dimensions. Academic research has been mostly concerned with expressiveness of service descriptions, while industry has focused on modularization of service layers for usability in the short term (Casey, Pahl 2003, Khilwani, et al. 2009). A significant amount of joint venture research is still required by academia and industry to achieve semantic environments in the enterprise community.

Considering the importance of virtual enterprise collaboration in the current global market and significance of information sharing and knowledge management in a collaborative environment, this research focuses on developing semantic frameworks for understanding and managing enterprise information resources. The key research gaps identified in different stages of virtual enterprise collaboration and addressed using the above semantic technologies in this research are as follows:

- *Identification*: Standard way of publishing enterprise information on the web.
- *Creation*: Automatic or semi-automatic approach for identifying suitable business opportunities and business partners for future collaboration.
- *Operation*: Annotating information resources with enterprise specific meanings to reduce the risk of misinterpreting the information.

- *Termination*: Methodologies to facilitate the meaningful acquisition, sharing and reuse of enterprise information.

This research work aims to exploit the emerging semantic technologies e.g. ontology modelling, semantic web services and information retrieval techniques to address the above research gaps and facilitate the process of enterprise collaboration and automated integration.

C h a p t e r   3

SCOPE AND OBJECTIVE OF RESEARCH

If I have seen farther, it is by standing on the shoulders of giants.
Sir Isaacs Newton

ABSTRACT

The paradigm of virtual enterprise is a predominant area of research and technological development for today's progressive industries. The aim of my research is to identify information sharing and knowledge management problems in each stage of virtual enterprise life cycle and subsequently propose solutions to resolve those problems.

CHAPTER OUTLINE

3.1.   Background

3.2.   Research Aims and Objectives

3.3.   Research Scope and Methodology

3.4.   Thesis Outline

3.1. Background

The competitive and frequently changing business and technical environments have made it essential for enterprises to enhance productivity, shorten product life cycle, increase product customization, and improve responsiveness to remain viable in the market. To achieve these seemingly conflicting objectives, enterprises are readily employing computing and other technical intelligence to provide them a consistent route towards continuous improvement. In the current global economy, the key to success is the extent to which business intelligence is utilized, and ideas and innovations are embedded in all sectors of the economy (Robinson 2004). The way a business gathers, shares and exploits knowledge is central to its ability to develop successfully. It is concerned with building an insight, foresight and knowledge about their internal management and effectively managing their present and shaping the future.

The role of knowledge in an economy is not a new idea, nor a new fact (Houghton, Sheehan 2000). All economies, however simple they are, place great importance on the generation and exploitation of knowledge and experience in the creation of wealth. However, the difference with this emerging economy is the magnitude of information and knowledge that is a variable for incorporation into economic activities, and this has changed the basis of competitive advantage (Fragidis, et al. 2008). In the past, syndication was less of a challenge because in a single enterprise everybody was aware of what others do. However in the current business climate, knowledge management does not occur naturally, when the market is evolving and enterprises are developing virtual collaborations to respond to business opportunities.

To achieve and exploit the benefits potentially available from knowledge sharing and communication, it is essential to identify, store, update, share, promote and transfer knowledge efficiently. The success of a virtual enterprise is primarily dependent on the effective exchange of accurate and meaningful information. Muata et al. identified five fundamental characteristics for the success of virtual collaboration, i.e. effective

coordination of industrial activities, highly interactive systems, external forces for business environments, internal as well as external interactions, effective control mechanisms for product/ process design and manufacture (Muata, et al. 2004).

3.2. Research Aims and Objectives

The concept of virtual enterprise is the most competitive approach and a predominant area of research and technological development for today's progressive industries. It consists of a balanced system composed of numerous autonomous entities (e.g. planning, design, procurement, production, inventory, marketing, distribution, sales and management) interrelated for producing usable products and services. In such collaborations, enterprises temporarily share competencies and resources to respond to the business opportunities in a more collaborative rather than competitive manner. Sharing and communication are often in response to opportunities or challenges that cannot be anticipated in advance and require a rapid response. Hence, it is very essential for virtual enterprise to share and exchange relevant information with other collaborators.

A virtual enterprise is created, managed, operated and monitored by the enterprises that are willing to collaborate and integrate their applications, processes and people. However, the success of virtual collaboration depends on knowledge management, i.e. on the effective acquisition, integration, sharing and management of information and knowledge resources in all stages of virtual enterprise. Information sharing is a basic requirement in a virtual collaboration, i.e. it is important to share information resources with the partners, so that the network can work as a single integrated unit. The information resources are converted into reusable and useful knowledge that is shared and managed with the people who need it. Hence through knowledge management, organizations seek to create an environment where information resources are used for improved performance, competitive advantage and continuous improvement of the enterprise or virtual collaboration.

Information sharing and knowledge management are therefore crucial challenges in an open networked virtual collaboration where many different and independent enterprises cooperate and share heterogeneous information resources. Accessing heterogeneous and distributed information resources in a coordinated and virtual way requires adequate semantic understanding to enable a seamless access and retrieval of the right information resources, while preserving the information representation and management requirements of each partner involved in collaboration (Castano, et al. 2006). In order to support the information sharing and communication, it is essential that the information resources are represented in a way that is usable and understandable by the networked organizations.

Keeping the above requirements in mind, this research aims:

To analyse the life cycle of a virtual enterprise, identify problems and requirements for information sharing and knowledge management tasks in each stage of virtual enterprise and propose semantic web framework and related web services to facilitate those tasks.

Corresponding to the main research aim, the overall objective of this research is to propose tools and techniques for providing explicit semantic description to all information resources shared and exchanged among heterogeneous enterprises in different stage of virtual collaboration. The objective is decomposed into the following four sub-objectives which are to be achieved step by step during the development of the work.

1. To analyse the life cycle of virtual enterprise and identify information sharing and knowledge management issues in each stage

2. To explore the need of explicit semantic description and machine understanding in information sharing and knowledge management at each stage of enterprise collaboration.

3. To develop a framework to facilitate enterprise collaboration with emerging technologies, e.g. semantic web, ontology and information retrieval techniques.

4. To evaluate the quality and feasibility of the proposed framework on an experimental and real life case study.

The next section provides a brief description of the research approach adopted to meet these objectives and the methodology applied in this research.

3.3. Research Approach and Methodology

As discussed in Section 2.5, this research focuses on information sharing and knowledge management issues of the virtual enterprise as identified in Table 2-3 by analyzing different styles of virtual collaboration. In particular it addresses research gaps discussed in section 2.7. The life cycle of virtual enterprise designed to explore these issues is mainly divided into 4 stages, as shown in Figure 2-2, namely- Identification, Creation, Operation and Termination. An initial, detailed review on life cycle of virtual enterprise, problems identified in each stage and research carried out in this thesis is briefly discussed as follows:

1. *Identification Stage*: This is the stage where enterprises come into the open market looking for new windows of opportunities. To explore opportunities in the global market, enterprises publish their skills, capability, expertise and competence that they are able to offer. Awareness of these profiles is a beneficial step to envisage and plan execution of new missions, to make a better appraisal of human capital, to initiate and mediate partnerships and to enhance organizational performance (Canavesio, Martinez 2007). Currently, enterprise competence profiles are usually captured and published as unstructured text documents, such as web homepages, industrial profiles, business cards etc (Linthicum 1999). There is therefore a need to develop a standard mechanism for modelling, management and exchange of competence information in a more explicit and structured manner.

    Chapter 4 explains the solution designed to solve this problem, i.e. *ECOS* (Enterprise Competence Organization Schema) ontology, which is introduced to represent   the   published   competences   of   enterprises

(http://kmm.lboro.ac.uk/ontologies/ecos/). Ontology based knowledge representation is a powerful scheme to model concepts and relationships that allows knowledge to be designed, shared and reused in a disciplined way (Fox 1993). *ECOS* ontology provides a metadata for annotating enterprise information in an electronic business card of a company and expresses it as a semantic resource which can be understood by systems as well as humans. The enterprise information is provided in two formats, i.e. human readable (HTML/XML) and machine readable (RDF/XML). The objective of *ECOS* is to create a web of machine readable pages with sets of interconnected data and semantic models that communicate and exchange knowledge and enhance the opportunity of enterprise collaboration.

2. *Creation Stage*: This is the stage where prospective enterprises develop virtual collaborations to extend their business potentials and respond to the new business opportunities. Generally, a broker is employed as a driving force for maintaining a collaboration pool of potential enterprises, searching and recognizing market opportunities, planning and drafting the collaboration (that can eventually become the virtual enterprise), and designing the project infrastructure. To accomplish these functionalities, the broker must have knowledge about the partners willing to develop networks. A collaboration registry for enterprise is needed and a database of enterprise knowledge must be maintained, so that the broker can identify new business opportunities and create awareness about the opportunities among pool partners.

In an attempt to support the broker, a platform is developed in the creation stage to map and measure relationships among enterprise profiles. The platform is structured in three main modules, i.e., ECOS-web – to maintain the collaboration pool, ECOS-track – to track business opportunities and ECOS-match – to categorize the pool and create awareness in the pool. Enterprises looking for new business opportunities can use *ECOS* to create a semantic profile and publish it on the web. The *ECOS-web* module analyzes and categorizes *ECOS* profiles based on their

competences and provides a visual and mathematical analysis of enterprise relations, ranging from general links to unidentifiable relationships. The ECOS-track is used to match capabilities and skills of enterprises with the requirements of business opportunities. Finally, the ECOS-match module compares the ECOS profiles of different enterprises to fulfil the requirements of a business opportunity and propose potential virtual collaborations. Details of the design of the platform and these modules are provided in Chapter 5. In Chapter 8, a prototype web application is developed in a Java based environment to represent the real life implementation of the above three modules.

3. *Operation Stage*: A major task in the operation phase is to deliver *the right information to the right person at the right time*. However, compared to single enterprises, knowledge sharing and communication among multidisciplinary teams is a notoriously challenging and complex problem (Huang, et al. 2005). Information resources are inherently distributed and often available in different formats e.g. electronic documents, databases, hardcopy, etc Amalgamating such information for sharing and exploiting across multiple business functions and processes clearly has attraction since it reduces the potentials for errors caused by data duplication and the resulting inconsistencies. But it is difficult to accumulate, maintain and modify heterogeneous knowledge in business processes which require changes frequently.

Bearing in mind the present requirement, a specialized knowledge service is proposed in this research that gleans information from documents and converts it into a semantic web resource using RDF and RDF Schema. The proposed text mining application, TEXT2RDF, provides explicit semantic descriptions to the enterprise documents and encodes them in an unambiguous and machine understandable format. This application is developed using the Stanford NLP parser to provide part-of-speech (POS) tags to the text content and then a Java based web application is developed in this research to extract terms and phrases from the text content. Subsequently, the keywords and phrases extracted from the documents are

annotated with domain specific terminology. The primary benefit of this new technique is to organize and share information resources in a unified and explicit manner, without any risk of misinterpretation (Refer to Chapter 6).

4. *Termination Stage*: Finally at some point of time, goals of virtual enterprise are achieved and partners are disbanded from the project. At this stage, the project related information is archived and distributed among partners; to learn lessons and utilize it in future projects. Such knowledge resources are a primary wealth for enterprises and can be utilized in enhancing organizational performance. However to achieve and exploit the benefits available from past knowledge, it is essential to store, update, share, promote and transfer data and information efficiently. This observation stresses the requirement for employing knowledge intensive methodologies that facilitate the meaningful acquisition, sharing and reuse of information.

   Knowledge management techniques are employed in enterprises that take organization's data and information and turn it into knowledge for tangible benefits for the company. Corporate memory, an explicit and disembodied representation of knowledge, is maintained to capture collective knowledge possessed by human capital and technology, past projects and present work, organizational structures and organizational culture; and makes it available to the entire company. This research work aims to provide explicit semantic descriptions of the information available in corporate memory and encode it in an unambiguous and machine understandable form. Initially, the proposed TEXT2RDF technique is used to convert a plain document into a semantic document. Subsequently, the semantic text is analyzed to identify the relationships between phrases present within a document and among other semantic documents available in corporate memory (Reference to Chapter 7). The relationships are identified using an indexing and retrieval method, Latent Semantic Analysis (LSA), which is capable of identifying patterns of relationships among a collection of documents.

The overall objectives and contributions of this research are provided in Figure 3-1. The semantic web framework proposed for each stage of the virtual enterprise are analyzed and evaluated as a case study. An experiment study with 25 enterprises has been prepared to illustrate the applicability and effectiveness of technologies proposed in this research. The prototype web application is developed in this research for the above case study and and implemented in a JAVA based environment. The research methodology adopted to formulate the above mentioned approach and apply in each stage of virtual enterprise life cycle is as follows:

1. *State-of-the-art survey*: This step includes the investigation and analysis of the life cycle of a virtual enterprise and detailed survey of existing methods, tools and techniques proposed to facilitate the virtual enterprise collaboration.

2. *Analysis of the requirements*: In order to focus the research area to be addressed, this step identifies the problems associated with information sharing and knowledge management and determines the requirements and possible solutions for this research.

3. *Development of the approach*: This step includes the development of a semantic web framework for assimilating the heterogeneous information resources by providing them with explicit semantic descriptions and encoding them in an unambiguous and machine understandable form.

4. *Prototype application*: This step includes implementation and testing of the prototypical environment of the proposed framework on a real life case study.

Following the above research approach and methodology, the research has been deployed and it has resulted in two major contributions:

IDENTIFICATION	CREATION	OPERATION	TERMINATION

OBJECTIVES

- Maintain collaboration pool
- Prepare member joining form

- Identify new business opportunities
- Search partners and suggest collaboration

- Support knowledge sharing
- Evaluate project, people and task

- Maintain corporate knowledge database
- Learning lessons from project

CONTRIBUTIONS

- **ECOS**: Ontology for representing competences
- **ECOS-form**: Web application for creating ECOS profile

- **ECOS-web**: Network Analysis and Visualization
- **ECOS-track** and **ECOS-match**: Selecting tenders and partners

- **TEXT2RDF**: Creating semantic web document
- **Hyperlinking keywords** with domain knowledge

- **Managing Corporate Memory** on semantic web
- **Identifying relations** among corporate documents

Figure 3-1: Objectives and contributions of this research

1. ECOS Ontology (Enterprise Competence Organization Scheme): An ontology proposed to capture the enterprise competence in a computer understandable manner. In this research, ECOS is used in first two stages of virtual enterprise life cycle, i.e.

   - Identification stage: To capture the basic information and competences of enterprises e.g. e.g. address, contact person, skills, capabilities, assets, plans etc.

   - Creation stage: ECOS is utilized to identify direct and indirect relations among enterprises and select appropriate business opportunities and partners.

2. TEXT2RDF Application: A text mining approach to extract basic information (e.g. terms, keywords, phrases, important phrases, sentiments etc) from a document and convert it into a computer understandable resource using semantic web standards. This application is utilized in last two stage of life cycle, i.e.:

   - Operation stage: To extract keywords and phrases from documents shared between enterprises and annotate with the domain specific meaning to reduce the risk of misinterpretation.

   - Termination stage: To identify relationships between terms and phrases present within a document and among other semantic documents available in enterprise information repository.

3.4. Thesis Outline

This thesis is divided in four parts and a brief description of each part is as follows:

1. *Background*: Chapter 1 sets the context for this research and identifies the importance of enterprise collaboration in current market environment. In Chapter 2, a detailed overview of virtual enterprise collaboration and the role of information sharing and knowledge management are provided. Chapter 3 lists the objectives of this research, provided an overview of the approach and presented the research methodology for this research.

2. *Theoretical research*: Chapter 4- Chapter 7 provide the theoretical and methodology contribution of this research for each stage of virtual enterprise.

3. *Experimental research*: In Chapter 8, an experimental study is provided to test the implementation of the proposed frameworks on real enterprise data.

4. *Conclusion and future work*: Chapter 9 evaluates the proposed research, discusses its novelty, summarizes the contribution, and concludes the thesis with recommendations for future work.

C h a p t e r   4

IDENTIFICATION STAGE: CAPTURING PUBLISHED COMPETENCES

Competence, like truth, beauty and contact lenses, is in the eye of the beholder
Raymond Hull

ABSTRACT

During the identification stage, capturing and managing competencies acquired, required and desired by an enterprise and further representing this information in a structured manner is a beneficial step for enhancing enterprise performance. Ontology is emerging as an effective tool to structure competences for comprehensive and transportable machine understanding. In this chapter, the ECOS ontology (Enterprise Competence Organization Schema) is presented as a mechanism to capture enterprise competence in a computer understandable manner.

CHAPTER OUTLINE

4.1. Introduction

With the advent of information and communication technologies, numerous formal and informal networks are emerging among organizations and individuals (Cross, et al. 2007). Companies are coming out of their self-centred environment to the open market for sharing and effectively utilizing their skills, core competences and resources. To explore opportunities in the global market, enterprises generally publish their skills, capability, expertise and competence that they are able to offer. Generally, competence models are used to consolidate and contextualize enterprise skills and capabilities in an explicit and structured manner (Sanchez 2004, Schoemaker 1992). A competence model is a crucial aspect in business scenarios that gives a deeper description about the profile of an enterprise or individual (Li, et al. 2000).

A competence profile demonstrates the knowledge, skills, experience, and attributes of an individual, group or organization necessary to carry out a defined function effectively. It includes their business activities, products and services produced and their preferences. In the past, competence was handled by humans having good knowledge about capabilities and requirements of the firms (Boucher, et al. 2007). However in the current situation, where global markets and partnerships are increasing, it is really difficult to handle information about all the potentially collaborative firms and hence it is necessary to model knowledge in a form that is manageable and addressable (Chiesa, Manzini 1997).

As discussed in Chapter 2.6, ontology is emerging an effective tool to structure enterprise information for comprehensive and transportable machine understanding (Goossenaerts, Pelletier 2002, Sugumaran, Storey 2002). In this research, the ECOS (Enterprise Competence Organization Schema) ontology has been designed and implemented to represent enterprise information (e.g. products, services, skills etc) and establish common vocabularies which can be easily interpreted by the system as well as humans. The ECOS ontology is proposed for expressing and sharing enterprise

competences, defining common vocabularies for similar concepts and supporting intelligent queries over heterogeneous databases. It captures enterprise published competences (information that company is willing to make public) with a consistent and comprehensive set of concepts and vocabulary and converts it into a semantic web resource using the Web Ontology Language (OWL) (Patel-Schneider, Horrocks 2004, Corcho, et al. 2003).

ECOS is presented as a mechanism to capture explicit enterprise competence knowledge in a computer understandable manner. The objective behind ECOS is:

To create a web of machine readable pages describing basic information and competences of enterprises with sets of interconnected data and semantic models.

The ECOS ontology organizes enterprise competences under four fundamental domains i.e. *General Information*, *Business Details, Specific Information* and *Business Records*. This ontology is used to create ECOS-card, a concept of electronic business card proposed in this research to publish enterprise competence in a machine readable as well as human readable manner. This enhances the manageability of enterprise competence information (which is generally only available as unstructured text documents) and enables software agents to carry out sophisticated tasks e.g. planning, evaluation, profile matching, need analysis, etc. A simple web application, named ECOS-form, has been developed in this research to create an ECOS-card and publish it on the web.

The novelty of this research resides in providing an ontology modelling technique to represent the enterprise competence and publish it in the form of an ECOS-card, which is a machine readable page in RDF/XML format. In this research, ECOS is utilized to capture and store the profiles of the enterprises that are willing to publish their competences and enter the market for networking and collaborations. Initially at the identification stage of a virtual enterprise, an ECOS-form is provided to the enterprises to enter their competences and an ECOS-card is created for them. The ECOS-card is used to support

enterprises in constructing and maintaining their own competences, evaluating their competences, matching their own competences with others and finding others with desired competences in appropriate domains. Further, this ECOS-card is also utilized in several ways in the creation stage of a virtual enterprise, such as in selecting suitable partners, identifying appropriate requests for tenders or other possible business opportunities, finding certain competencies, etc.

4.2. Enterprise Competence

Any company willing to enter a global market must publish their competences e.g. skill, capability, expertise, etc that they are able to offer. When enterprises are considered, competences include skills and abilities of an organization to carry out certain types of tasks based on knowledge and experience of their methods and resources. For example, a company XYZ designs and develops IT software and provides support for it, then the software is its product or service and IT support for their software is its skill. Competence is the ability to coordinate the deployment of a company's assets in ways that give competitive advantage and produce success in the market place. Formally, enterprise competence is defined as (Allen 2003):

> A specific, identifiable, definable, and measurable knowledge, skill, ability and/or other deployment-related characteristic (e.g. attitude, behaviour, physical ability) which a resource may possess and which is necessary for performing an activity within a specific business context

Competence is a new way to consider relationships among individuals, groups and organizations, associated with a new vision of performance as well as new forms of management and decision making (Prahalad, Hamel 1990). Understanding and sharing competences improves performance of firms in the following aspects (Jussupova-Mariethoz, Probst 2007):

- attract, retain and improve the best available resources for creating and realizing continuous value creation and distribution,

- publish the competence of one's own firm in the market and identify potential cooperation opportunities,

- increase awareness about one's own current capabilities as well as understand competences that other companies can offer, which can be used to identify areas for future development, and

- initiate or mediate new partnerships.

4.2.1. *Related Literature*

Modelling and management of enterprise competence refers to how information is managed with respect to organizations, groups and individuals (Vernadat 2002). Specialized knowledge services are used to extract knowledge and represent them systematically in a formalized manner (Sanchez 2004). Ontology modelling is one of the tools that provide an explicit description of real world knowledge (Gruber 1993). The term ontology has been coined from philosophy, as a branch of meta-physics concerned with the systematic explanation of existence. Due to its strong implications for conception of reality, it has gained much interest in artificial intelligence for defining the basic terms and relationships comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary. It is utilized in knowledge representation, knowledge engineering, information modelling, data base development and integration, information retrieval and extraction, etc (Gruber 1993, Huang, Diao 2007, Brandt, et al. 2008, Patil, et al. 2005). It basically contains concepts, concept taxonomies and relationships to support information exchange and knowledge sharing.

There is a plethora of literature related to the development of ontology models (Huang, Diao 2007, Madni, et al. 2001, Uschold, Jasper 1999). But only a handful of efforts can be seen that have attempted to model enterprise competence related information in a formal ontology. However, other work has been done to make this modelling process easier or has modified it for a particular purpose. Without any

pretention of providing an exhaustive review of the published literature, the most relevant contributions in the area of enterprise modelling and competence modelling are summarized in next two sections respectively.

4.2.1.1. *Enterprise Ontologies and Modelling Framework*

Enterprise ontology models are generally used for enterprise knowledge management with well defined terminology accepted by all participating firms. The goal of an enterprise ontology model is to define semantics for the domain terminology in an easy to understand way. Such an ontology model is similar to a dictionary, taxonomy or glossary, but with structures and formalisms that enable computers to process its contents. An enterprise ontology mainly consists of terms and contents for (Schlenoff, et al. 2000, Gottgtroy, et al. 2006):

- business entities and relationships of enterprise,
- processes and planning,
- structure of organization,
- market details and products/services,
- high level planning and preferences of enterprise.

Some important enterprise ontologies developed by the artificial intelligence community and enterprise modelling community are (Madni, et al. 2001, Khilwani, et al. 2009):

1. *TOVE* (TOronto Virtual Enterprise): This is developed to support enterprise integration with minimum ambiguity and maximum understanding and precision in communication (Fadel, et al. 1994). TOVE provides a generic and reusable ontology for modelling enterprises. It is basically the group of ontologies comprising of three different categories of ontologies viz. core ontologies for products and services; derivative ontologies for transportation and inventory; and an enterprise ontology for enterprise design and business processes.

2. *Enterprise Ontology*: The Enterprise Ontology is developed under the Enterprise project to provide a collection of terms and definitions relevant to business enterprises (Goossenaerts, Pelletier 2002). The Enterprise Ontology was not developed for a specific type of enterprise, but as a generic model oriented towards business and organization rather than towards a specific domain. This ontology has five top-level classes for integrating the various aspects of an enterprise

   - *Meta-Ontology and Time* – terms used to define the terms of the Ontology (e.g. Entity, Relationship, Role) and terms related to time (e.g. Time-Interval)

   - *Activity, Plan, Capability and Resource* – terms related to processes and planning (e.g. Activity, Planning, Authority, Resource, Allocation)

   - *Organisation* – terms related to how organisations are structured (e.g. Person, Legal Entity, Organisation Unit)

   - *Strategy* – terms related to high level planning for an enterprise (e.g. Purpose, Mission, Decision, Critical Success Factors)

   - *Marketing* – terms related to marketing and selling goods and services (e.g. Sale, Customer, Price, Brand, Promotion)

3. *OIM* (Open Information Model): The goal of the OIM project, sponsored by Meta Data Coalition (MDC), is to provide a vendor-neutral and technology-independent specification of enterprise meta-data(Meta Data Coalition 1999). The major components of OIM are: Analysis and Design Models, Objects and Components Models, Database and Data Warehousing Models, Knowledge Management Models, and Business Engineering Models. The major concepts defined in the Business Engineering component deal with the organizational aspect of enterprise models.

4. *IDEON^{TM}*: A unified enterprise ontology that provides a common foundation for designing, reinventing, managing and controlling collaborative, distributed enterprises (Madni, et al. 2001). It consists of a set of business entities and relations that represent common entities with an enterprise context. IDEON employed four

complementary perspectives to capture the key concepts and relationships that characterize an enterprise. These perspectives are as follows:

- *Context view*: To represent interaction between an enterprise and its external environment.

- *Organizational view*: Structural view of enterprise to complement the enterprise context view.

- *Process view*: To represent plans, processes and activities of the enterprises.

- *Resource/Product view*: To elaborate different types of resources, such as material resource and human resource, needed to execute the processes.

5. *BPML* (Business Process Modelling language) represents a new family of process definition languages intended for expressing abstract and executable processes that address all aspects of enterprise business processes (Arkin 2002). It provides an abstract execution model for collaborative and transactional business processes based on the concept of a transactional finite state machine.

6. *CNO* (Collaborative Network Organization): In order to share the same terminology, the Virtual Enterprise Partners need to agree on the terms that they intend to use in collaboration. CNO is introduced to structure concepts and terms used in the field of collaborative networks (Plisson, et al. 2007).

A detailed survey on each of these ontologies and other relevant enterprise ontologies and modelling frameworks is provided in my review paper, **Appendix A** (Khilwani, et al. 2009).

4.2.1.2. *Competence Modelling*

Competence modelling and management presents specific requirements in enterprise modelling. Enterprise competence models are generally used for representing relevant business activities and products or services, offered or required by the company. Several competence management tools have been explored during the last few years. Lv and Zhu proposed an intelligent recruitment framework using recruitment websites which can

automatically find the appropriate job-seekers for a certain post and vice-versa. A skill based graphical ontology model is developed in (Hexin Lv, Bin Zhu 2006) for calculating the similarity between acquired individual competence and industrial requirements. Yu et al. developed an Enterprise Competence Model (ECM) for structural representation and exchange of competence within and outside the enterprise (Li, et al. 2000).

Pareto and Snis proposed a flexible interactive visualization model to support competence management from the perspectives of organizations and groups as well as individuals (Pareto, and Snis 2007). Harzallah et al. built a CRAI (Competence-Resource-Aspect-Individual) competence model which is useful for different competence management processes, e.g. enterprise competence identification, personnel recruitment, etc (Harzallah, et al. 2006). According to Schmidt and Kunzmann, an enterprise uses competence in two different perspectives, i.e. at the organizational level for integrating human assets into management practice and at the operational level for fostering learning activities of individual employees (Schmidt, Kunzmann 2006). They developed a reference ontology to bring the above two perspectives together. Draganidis et al. developed an ontology based tool for competence management (Draganidis, et al. 2006).

Tarassov et al. modelled an individual and industrial competence model using ontology as a meta-model (Tarassov, et al. 2006). They utilized ontology as a way to formalize individual and industrial competence. Boucher et al. has done a general survey on the formalisation and integration of competence oriented concepts within enterprise information systems and decision systems (Boucher, et al. 2007). Pepiot et al. proposed UECML (Unified Enterprise Competence Modelling Language) with the intension of providing a neutral interface for enterprise competence modelling (Pépiot, et al. 2007). This language is developed to alleviate the deficiencies of UEML (Unified Enterprise Modelling Language) and make a simple and single model for competence modelling. Jussupova-Mariethoz and Probst proposed the Business Concept Ontology (BusCO) modelling concepts related to the enterprise performance and competence monitoring

processes (Jussupova-Mariethoz, Probst 2007). BusCO gives a common understanding of the data and information which facilitates the analysis of the general and enterprise specific concepts related to competences and experience monitoring.

4.2.2. *Need for A Published Competence Model*

To fully exploit enterprise skills and capabilities in a global market, it is necessary to consolidate and contextualize published competences in an explicit and structured manner. Published competence is a very crucial element for a company wishing to participate in collaboration (Sanchez 2004). A published competence consists of knowledge that the enterprise is willing to make public (or publish). It includes details that the company wants to tell others about who it is / what it can do / what it is interested in / what it wants to achieve etc. It emphasizes information that may affect relationships with individuals, groups and organizations, associated with a new vision of performance as well as new forms of management and decision making. Unfortunately, most of the models discussed above do not cover this requirement and are unable to represent entire competences with the same, single formalism (Jussupova-Mariethoz, Probst 2007, Draganidis, et al. 2006, Pépiot, et al. 2007). In the context of this research, the main deficiencies identified in existing literature and addressed in this work are:

- Enterprise ontologies are too general and mainly focus on single enterprise modelling – hence are not appropriate for enterprise collaboration.
- Competence models lack a consistent set of concepts and vocabulary for describing competences – hence they are not suitable for semantic web based information systems.

The ECOS (Enterprise Competence Organization Schema) is therefore proposed in this research in response to the incompleteness of existing work. The proposed ontology is based on the representation of published competences of enterprises in a formalized notation.

4.3. ECOS: An Ontology for Published Competences

The Web is a basic entity for making information electronically available so that enterprises can access and share resources and expertises over the internet. This technology supports information sharing and communication by providing platform independence to publish and access data. However, the current web technology arranges the information syntactically, assuming it to be semantically homogenous which commonly causes problems and misunderstandings (Xu, Xiao 2007). Enterprises often use different terminologies to describe the same meaning or alternatively the same terminology may be associated with different meanings. Such lack of clear semantic description of the meaning of contents in a particular domain has hampered the interoperability and sharability of knowledge between companies. In order to handle this interoperability challenge for future organizations, it is necessary to make knowledge explicit and machine interpretable (Vernadat 2002, Pépiot, et al. 2007).

The next generation enterprise community needs sets of interconnected data and semantic models to communicate and exchange their knowledge, without the current risk of misinterpretation. The Semantic Web is a vision for extending current web technology in which information is annexed with a well defined meaning that enhances the interoperability of computers and people (Casey, Pahl 2003). This extension of the web employs ontologies for defining semantics and thus gives meaning to the data and applications for automatic processing. The key goal of the semantic web is to enable the access of data from distributed locations and use ontologies to aggregate and utilize it. Ontologies are considered to be the basic building blocks of the semantic web with a collection of concepts and relations to provide machine supported data interpretation and reduce human involvement in data and process integration.

An ontology aims to capture the semantics and relations of a domain, make them explicit and eventually code them in symbolic systems so that they can be manipulated and exchanged. A significant effort has been devoted by the research community to

develop ontology representation formalisms that can be shared and processed automatically using scripts or programs. The first ontology introduced for the semantic web was Friend-of-a-Friend (FOAF), a consistent set of vocabulary for describing a person's social sphere e.g. professional and personal lives, their friends and interests etc (Brickley, Miller 2007). Subsequently, various ontologies and micro-formats (such as, vCard, SIOC, SKOS, BIO, etc) provided explicit semantic descriptions for information resources and encoded them in a machine understandable format (Halpin, et al. 2006, Guarino 1998). In this research, the *ECOS* (Enterprise Competence Organization Schema) ontology is introduced to represent enterprise information and establish common vocabularies which can be easily interpreted by the system as well as humans.

ECOS is proposed for expressing and sharing enterprise competences, defining common vocabularies for similar concepts and supporting intelligent queries over heterogeneous databases. It captures enterprise published competences (information that an enterprise is willing to make public) with a consistent set of concepts and vocabulary and converts it into a semantic web resource. ECOS does not aim to propose a definitive enterprise modelling language acting as a substitute for all previous proposed ontologies; rather it provides a standard meta-model for published competences of enterprises. ECOS and its developments contribute towards:

- *Metadata for electronic business cards*: Metadata is machine processable information for the web, which when structured into a hierarchical arrangement is known as an ontology. Both terms describe "*what exists*" for some purpose and are used as a shared means of communication between computers and between humans and computers (Gruber 1993). To reach this goal, semantic annotations are attached to information so that they can be processed and used by computers. The ECOS ontology provides metadata for annotating information in an electronic business card and expressing it as a semantic resource.

- *Consistent set of concepts and vocabulary*: Standards are used to represent competences with a consistent set of vocabulary, which is widely accepted and

shared among enterprises. In an attempt to articulate a comprehensive vocabulary, classifications defined by United Nations Statistics Division and European Union are considered in this research, see **Appendix B**. These international classifications, including CPC (Central Product Classification), ISIC (International Standard for Industrial Classification), NUTS (Nomenclature of Territorial Units for Statistics), CPV (Common Procurement Vocabulary) and ISCO (International Standard Classification of Occupations), are converted into semantic web resources using SKOS (Simple Knowledge Organization Scheme (Miles, et al. 2005).

- *Interoperability with existing ontologies*: In order to avoid reinvention of the wheel, existing ontologies are used for explaining certain ECOS concepts. The vCard ontology is used for capturing information about person and address in the ECOS ontology, see **Appendix C** (Iannella 2001).

4.3.1. *ECOS: Definition and Concepts*

Apriori, to understand the implementation of ontologies in competence management, it is necessary to understand what they are. In recent times, ontologies have gained increasing interest from AI communities due to their capability to capture knowledge for representation, processing and management (Linthicum 1999, Browne, Zhang 1999, Ratchev, et al. 2000). In situ, ontologies have been defined by several researchers in different ways but the definition that is most general and widely quoted is "*an explicit and formal specification of a shared conceptualization*", where explicit means it comprises concepts, properties, relationships, functions, axioms and constraints (Gruber 1993, Guarino 1998, Guarino, Giaretta 1995). Similarly, formal means it is machine readable and interpretable. Conceptualization is used as an ontology provides an abstract model which gives a simplified view of the things it represents and shared means there is consensus about the information which is acknowledged by a group of experts. In short, it is a set of concepts, its taxonomy, interrelations and the rules that govern such concepts.
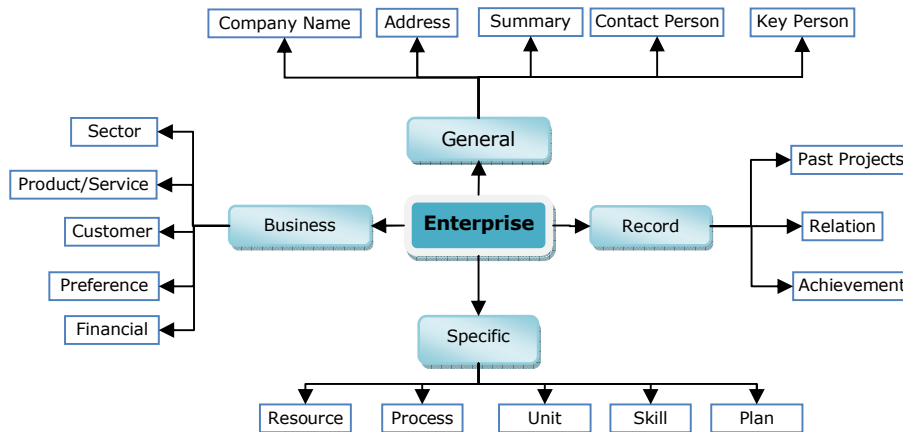
Figure 4-1: ECOS Ontology Model

The ECOS ontology proposed in the research for modelling enterprise competence is shown in Figure 4-1. It is mapped as a generic tree structure to model the published competences of enterprises. The structure has been divided into 4 key classes, namely - General Information, Business Details, Specific Information and Records. These classes are further divided into subclasses to represent common entities used to represent enterprise competences. The design of ECOS was inspired by the recognition that enterprise ontology models can make processing and sharing enterprise information more straightforward, systematic, effective and much simpler. The existing ontology models, mainly Enterprise Ontology with five top level classes and IDEON[TM] with four complementary perspectives (discussed in Section 4.2.1.1), were amended for developing the ECOS ontology model shown in Figure 4-1.

The 7 stage ontology design process proposed by Noy and McGuniness has been followed for developing the ECOS ontology (Noy, McGuniness 2001). The knowledge engineering approach includes the following steps:

1. *Determine domain and scope of ontology*: The domain and scope of ECOS is clear from the beginning. It is used to model the published competences of enterprises.

2. *Consider reusing existing ontology*: In order to ensure interoperability of the ECOS ontology with other ontologies and to avoid reinvention of the wheel, existing ontologies (vCard ontology) and standards are included in the proposed ontology.

   - vCard ontology is used for expressing name, address information, phone etc of companies and individuals. The concepts and properties of vCard utilized in the ECOS ontology are given in **Appendix C**.
   - Standard Classifications (given by the Statistical division of the UN) are used for developing a consistent set of vocabularies for enterprise competences (details in Appendix B). CPC codes (Central Product Classification) are used for products, tangible resources, etc; ISIC (International standard for industrial classification) for defining sector, skills, etc; ISCO (International Standard Classification of Occupations) for defining job roles.

3. *Enumerate terms*: The key concepts and relationships existing in the competence description have been identified and used for developing this structure. The model is developed with a consistent set of vocabulary and semantic definitions for concepts and relations. The terminologies used in this vocabulary set are defined using knowledge and experience of published information models on enterprise ontologies and competence models.

4. *Define class and class hierarchy*: The structure has been divided into 4 key classes, namely- General Information, Business Details, Specific Information and Records. These key classes are further divided into 25 classes in total.

5. *Define properties of classes*: The classes defined in the ECOS ontology are linked with a list of properties, 15 in total.

6. *Define facets of slots*: This step is covered in greater detail in the following sections.

7. *Create instances*: Instances of the ECOS ontology are shown in Chapter 8.

ECOS is used to capture and publish enterprise competence in a machine as well as human understandable manner. It is based on a set of core concepts and sub-concepts required to describe enterprise competence. ECOS is based on four high level concepts

that collectively represent the published competence of an enterprise. The detailed description of ECOS concepts are discussed as follows:

4.3.1.1. General Information

The ECOS representation starts with *General Information*, capturing basic information about the company such as name, address, contact person etc. General Information consists of 5 concepts:

- *Company Name (GCN):* Includes the name of the enterprise added with a URI (Unique Resource Identifier), such that:

$$GCN = \langle name, URL \rangle \qquad \text{4-1}$$

  $name$ ¡s the name of company. URL is basically a web link that leads to a web page (e.g. homepage) on the World Wide Web (WWW). It is used to give a unique name to each enterprise in the database.

- *Summary (GS):* Stores a brief description about the enterprise. This concept is used to capture a concise opening paragraph about the company.

$$GS = \langle summary \rangle \qquad \text{4-2}$$

- *Address (GA):* This concept expresses the contact address of the enterprise, including address, phone numbers, etc. The vCard ontology is used to express an enterprise's address's in the ECOS model, such that:

$$GA = \langle \{vCard\}, GA_{code} \rangle \qquad \text{4-3}$$

  where, GA_{code} is the NUTS code referring to the region of the enterprise.

- *Contact Person (GCP):* Terms defined in the vCard ontology are used to describe information about the contact person of the company.

$$GCP = \langle \{vCard\}, GCP_{code} \rangle \qquad \text{4-4}$$

  where, GCP_{code} is the ISCO (International Standard of Occupation) code selected for the contact person.

- *Key Person (GKP):* This concept is used to give information about key people responsible for managing the business and affairs of the company and ensuring its success. It is also expressed using the vCard ontology and ISCO codes.

$$GKP = \langle \{vCard\}, GKP_{code} \rangle \qquad\qquad \text{4-5}$$

This concept may include information about lead persons in the company such as shareholder, owner, director, etc. And GKP_{code} is expressed as an ISCO code representing the role of a person.

4.3.1.2. Business Details

The intention of this key class is to capture the market details of a company. It includes information about sector, product/services, customers, preferences and financial details of the company.

- *Sector (BS):* This concept refers to a broad way of grouping similar industries. Generally, enterprises are classified in four key sectors: the *primary* sector, largely raw material extraction industries such as mining and farming; the *secondary* sector, involving refining, construction, and manufacturing; the *tertiary* sector, which deals with services (such as law and medicine) and distribution of manufactured goods; and the *quaternary* sector, a relatively new type of knowledge industry focusing on technological research, design and development such as computer programming, and biochemistry. A fifth *quinary* sector has been proposed encompassing non-profit activities. In this research, Sector codes are used to classify companies into different sectors. Formally, the sector concept is defined as:

$$BS = \langle BS_{code} \rangle \qquad\qquad \text{4-6}$$

where, BS_{code} is the first level of ISIC code used for defining the enterprise sector (Appendix B).

- *Products/Services (BPS):* Generally, the description of product/service includes information about, what it is, what is does and its features and benefits. Keeping the above requirements in mind, this concept captures the description as:

$$BS = \langle name, BPS_{code}, detail \rangle \qquad \text{4-7}$$

where, $name$ refers to the formal name of the product, BPS_{code} is the list of CPC code (Central Product Classification) used for classifying product, $detail$ includes further information about the product defined under data type properties: price, description, specification, purpose, etc.

- *Customers (BC):* This concept lists the potential buyer or user of products/services of the company. This is typically through purchasing or renting of goods or services. Formally, it is defined as:

$$BC = \langle \{GI\}, \{BS\} \rangle \qquad \text{4-8}$$

where, GI is the general information about the customers, BS is used to define the sector code of the customers.

- *Financial Highlights (BFH):* The financial details of the company include information about their revenue, profit, income, employee etc. This information is captured under this concept, formally defined as:

$$BFH = \langle name, value, time \rangle \qquad \text{4-9}$$

where, $name$ is the type of information, $value$ is the value of $name$ and $time$ is the time when this information is captured.

- *Preferences (BP)*: Enterprises always look for new business opportunities in the market, but with a focus on certain objects of interest. This concept in included to capture those business interests of the company. It is formally defined as:

$$BP = \langle type, value, weight \rangle \qquad \text{4-10}$$

where, $type$ is the type of preference such that $type = \{sector, region, tender\}$; $value$ is the value of the preference selected from standard classifications, where,

$sector \in SectorCode, \quad region \in NUTS \quad and \quad tender \in CPV\ Codes; \quad weight$ represents the numerical weight of the enterprise preference.

4.3.1.3. Specific Information

The central idea for this key class is to capture core competences of an enterprise. Core competences are a set of tangible and intangible resources that are difficult to imitate or substitute and procure a competitive advantage for the enterprise. The concepts included in this class are:

- *Process (SP):* This concept defines the activities and programs that a company employs to meet its mission, objectives and interests. A process is basically a sequence of actions resulting in a product or service. In general, the description of process includes, process name, description, actors involved in the process, its start and finish time, sequence of activity, etc. However, in the ECOS model only a description of the information about published competence is included. Therefore, the formal description of SP includes:

$$SP = \langle name, SP_{code}, detail \rangle \qquad\qquad 4\text{-}11$$

  where, $name$ is the name of the process, SP_{code} is one or more ISIC codes defined for the process, and $detail$ is the description of the process.

- *Resource (SR):* Resources are entities used by processes in the day to day operation of the organization. This concept captures information of entities that support the execution of processes. Resources can be classified in two categories, material and non-material resources (Pépiot, et al. 2007). A resource (SR) can be formally defined as:

$$SR = \langle name, type, SR_{code}, detail \rangle \qquad\qquad 4\text{-}12$$

  where, $name$ is the name of the resource, $type$ is the type of resource such that $type = \{Material, Non-material\}$, SR_{code} is the CPC/ISIC code selected for the resource and $detail$ is the description of the resource.

- *Unit (SU):* This concept captures information about an organizational unit holding responsibility for identified processes and resources. A Unit is formally defined as :

$$SU = \langle name, SU_{code}, detail, actor, responsibility \rangle \qquad 4\text{-}13$$

where, $name$ is the name of the organizational unit, SU_{code} is the ISIC code identified for the SU, $detail$ is the details about the organizational unit, $responsibility$ is a set of processes and resources such that $responsibility = [\{SP\}\{SR\}]$ and $actor$ is the set of human resources used to support the handling of responsibilities.

- *Skill (SS):* It is the ability to handle responsibilities and perform actions as per the requirements. It is a repeatable pattern of actions executed to create, produce and/or offer product/services to the market. Skills arise from the coordinated activities of resources, processes in an organization unit. A skill is formally defined as:

$$SS = \langle name, SS_{code}, detail \rangle \qquad 4\text{-}14$$

where, $name$ is the name of the skill, SS_{code} is the ISIC code identified for skill and $detail$ is the description of skill

- *Plan (SPP):* Enterprises are always in search of new business opportunities and must be prepared to take on new challenges based on their capabilities and what they are able to perform, but which they may not carry out at present. This concept is used to capture information about what the enterprise believes they can perform in the future. Formally, the plan in defined as:

$$SPP = \langle name, SPP_{code}, detail \rangle \qquad 4\text{-}15$$

where, $name$ is the name of the plan, SPP_{code} is the selected ISIC code and $detail$ is the description of the SPP.

4.3.1.4. Business Records

Central to the Record Category is information that an enterprise publishes to capture the attention of customers and vendors offering business opportunities. This key class lists

information about past activities, relations and achievements of the company, etc. The concepts included in this class are:

- *Past Project (RPP):* This concept lists past projects of the company handled independently or in collaboration with other companies. Formally, it is defined as:

$$RPP = \langle name, RPP_{code}, detail, time, partner \rangle \qquad \text{4-16}$$

where, $name$ is the name of the project, RPP_{code} is the list of selected ISIC/CPC codes for describing project, $detail$ is the description of the project, $time$ is for start and end time or period of the project, $partner$ is the list of partners involved in the project.

- *Relation (RR):* An enterprise always maintains and publishes relations with customers, investors and other companies (competitors or collaborators) to develop new connections in the global market. This concept captures information, formally defined as:

$$RR = \langle name, RR_{code}, type, detail \rangle \qquad \text{4-17}$$

where, $name$ is the name of the company, RR_{code} is the selected sector code, $type$ is the type of relations, i.e. $type = \{Investor, Customer, Supplier, Collaborator, Competitor\}$ and $detail$ is the description of the relation.

- *Achievements (RA):* Companies often publish information about their achievements, including patents, standards, publications, etc. In ECOS, this information is captured as:

$$RA = \langle name, RA_{code}, type, detail \rangle \qquad \text{4-18}$$

where, $name$ is the name of the achievement, $type$ is the type of achievements such that $type = \{Patent, Standard, Publication\}$, RA_{code} is the ISIC/CPC code selected for achievement and $detail$ is the details of the RA.

4.3.2. *Coding ECOS Ontology*

An ontology must be coded in some formal language to express the concepts in the domain in a manner that computers can understand and manipulate easily. Several ontology languages have been developed during the last few years and have been used in the context of the semantic web. These ontology representation languages are logic based, frame based or web based. A comparison between all the existing ontology languages is provided in (Jussupova-Mariethoz, Probst 2007, Gómez-Pérez, Suárez-Figueroa 2004). The languages are extended versions of XML syntax facilitating greater machine interpretability and semantic interoperability. Ontology exchange language (XOL), Simple HTML Ontology extension (SHOE) and Ontology mark-up language (OML) are the primary extensions of the XML syntax. However, other popular ontology languages are RDF (Resource Description Framework), RDF Schema (RDF-S) and OWL (Web Ontology Language) developed by World Wide Web consortium (W3C). These allow the specification of the semantics of data based on XML in a standardized and interoperable manner.

The Web ontology language, OWL is the most expressive language for specifying, publishing and sharing ontologies (Noy, McGuniness 2001, Yalan, et al. 2006). It facilitates greater machine interpretability of web content than that supported by XML, RDF, or any other languages proposed in the literature. It is the only standard web ontology language that provides formal semantics along with additional vocabulary. OWL mainly comes in three sub-languages, OWL Lite, OWL DL and OWL full, which are intended to be layered according to their increasing expressiveness (Noy, et al. 2001). In this research, OWL DL is used to model the ECOS ontology as this enables the encoding, exchange and reuse of the ECOS ontology.

OWL describes the structure of an ontology by using some basic building blocks i.e. classes, properties and individuals. *Classes* are used to group and describe individuals with similar properties (implicit or explicit). *Properties* are binary relations on individuals,

i.e. linking two individuals together or associating individuals with instances. In OWL, properties have multiple domains that connect a property with a class and multiple ranges to link the property with a class or instance. *Individual* represents instances of the described domain. The above three building blocks are also associated with relationship i.e. each *class* has a set of *property*, *class* has instance *individual* and *individual* has value for *property*.

The OWL schema developed for the ECOS ontology consists of 25 classes, 15 properties and 6 existing ontologies. Out of 25 classes in total, 5 are main classes used to describe general information, business details, specific information and business records of the *enterprise* class. The other 20 classes are auxiliary classes used for describing the main classes and their relationships. Table 1 explains classes, subclasses, properties and external ontologies utilized in the ECOS ontology. The schema for the EOCS ontology can be obtained on the web at: http://kmm.lboro.ac.uk/ontologies/ECOS.

4.3.3. *ECOS-card and ECOS-form*

The OWL schema for ECOS provides a standard vocabulary for representing enterprise information and publishing it on the web. The concept of an ECOS-card is proposed in this research to create a semantic profile of an enterprise using the ECOS vocabulary. It is a novel concept of publishing information in a machine readable as well as human readable manner. As shown in Figure 4-2, an enterprise profile is represented in two formats i.e. human readable (front) and machine readable (back). The human interpretable page is written in plain HTML/XML format so that it can be viewed in web browsers; whereas, RDF/XML format is used for representing the machine readable information. Computers can use this card to find information about a company or to search companies having certain competences. This can also be used to facilitate the process of search engines and other search applications (e.g. tender selection, partner search) to work semantically.
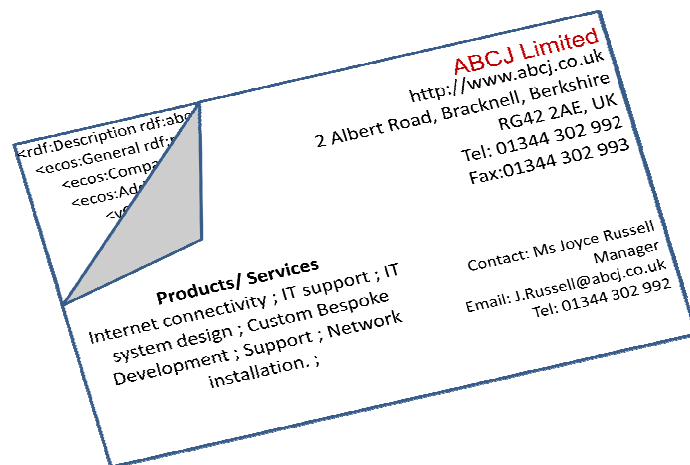
Figure 4-2: ECOS Card in two formats, Human readable in front and machine readable in back.

An ECOS-form is a simple web application developed to create an ECOS-card. It provides a platform for developing an ECOS-card in RDF/XML format. The ECOS-form is available on the web, http://kmm.lboro.ac.uk/ECOS.html. It is an application form with fields for all the sorts of information required for creating an ECOS-card. A classification list for standard classifications is also provided with this form to define and add codes with enterprise information.

4.3.4.   *Querying ECOS-card*

RDF/XML format, used for representing ECOS profile, is the most common approach for linking data with ontologies. It is a W3C standard format for storing arbitrary data on the semantic web. The data in RDF/XML format is described as instances of corresponding ontology represented as RDF schema or OWL ontology. The advantage of statement-centric RDF data format is that it can be stored in various formats. For example, graph based representation, which is especially suitable for humans because it is easily readable. Other forms of representation include XML syntax which is more verbose but more popular on the web; and N-3 or N-triple, which is less verbose and better readable, but rarely used and known. However, in this research RDF/XML syntax is used for storing ECOS profiles. An advantage of using an established syntax for ECOS is to gain existing

advantages of XML and have a variety of applications e.g. web publishing, web searching and automating web tasks, e-business applications, pervasive computing, etc.

For RDF/XML based syntax, the next requisite is a query language which can access the information stored in ECOS profiles. SPARQL (SPARQL Protocol and RDF Query Language) is the query language proposed for the semantic web and is currently a W3C recommendation. More precisely, SPARQL is a query language for RDF. In syntax, SPARQL resembles SQL (Sequential Query Language) to a great extent (with commands like SELECT, WHERE, ORDERBY) but the difference is SQL works on a relational database while SPARQL works on the RDF graphs.

4.4. Summary

During the identification stage of a virtual enterprise, the prerequisite for any enterprise is to publish their skills, capabilities, expertise and competences in the global market. In this chapter an enterprise competence ontology model, named ECOS, is presented as a mechanism to capture enterprise competence in a computer understandable manner. The ECOS ontology captures enterprise competences using a consistent and comprehensive set of concepts and vocabulary and converts them into a semantic web resource using the Web Ontology Language (OWL). ECOS is proposed with an idea of creating a web of machine readable pages describing companies, their basic information, skills, assets etc.

ECOS captures an enterprise's published competences (i.e. information that the enterprise is willing to make public) with a consistent set of concepts and vocabulary and presents and stores it as a semantic web resource. These semantic profiles are directly linked with standard classifications and existing ontologies, and are thus indirectly linked with each other. These hidden relations can be used for matching ECOS profiles and identifying relations between them. It can be used to support various enterprise applications related to the competences of an enterprise. Three such applications, i.e.

enterprise network analysis, tender selection and partner selection are shown in the creation stage of a virtual enterprise, in the next chapter.

The detailed example of how the ECOS ontology captures, presents and publishes competence in a formalized structure is provided in Chapter 8.

C h a p t e r   5

CREATION STAGE: EVOKING POTENTIAL COLLABORATIONS

Being in a band is always a compromise. Provided that the balance is good,
what you lose in compromise, you gain by collaboration
Mike Rutherford

ABSTRACT

During the creation stage of a virtual enterprise, a relatively *stable network of companies* needs to be developed to provide a basis for social relationships and trust among companies. To facilitate the creation of a virtual enterprise, a semantic web based platform is developed in this chapter that raises awareness about opportunities and enhances synergy among enterprises. The platform is structured in three main modules, i.e., ECOS-web – to maintain the collaboration pool, ECOS-track – to track business opportunities and ECOS-match – to categorize the pool and create awareness in the pool.

CHAPTER OUTLINE

5.1. Introduction

In the past, the key to success in industries was to invest exactly at the right place, at the right point of time, in order to be able to deliver products and services faster and cheaper than the competitors. Today things are no longer that straightforward. The current trend in the market is to develop collaborative networks to share skills, core competencies and resources for short or long term collaborations. A driver for the formation of a collaborative environment is the desire to develop a stable network of companies and effectively share costs, skills, and core competencies that collectively enable partners to access global markets. However, searching for appropriate projects and finding the right partners for the collaboration is a time consuming activity for an enterprise.

Enterprises in pursuit of emerging business opportunities publish their competences (for example, products, services, skills, plans etc) in the market or on the web to extend their business processes and develop collaborations with similarly focused companies. Often, companies register their profile with some tender alert services or may use their built-in software for getting new business opportunities. Such matching services list the business opportunities based on the resemblance of the company's profile with the project requirements. Without excluding notifications about any business opportunity, the alert services notify users of all relevant projects matching with the area of company's core expertise. However, most of business opportunities identified by alert services, upon examination fall low down the manager's priority list, as they would not provide a good business fit for the company. Also, companies neglect opportunities which they could not undertake it individually, and in this way, they also overlook some opportunities in which they could provide input in collaboration with other companies.

A collaboration registry therefore needs to be developed to maintain a database of enterprise knowledge for building stable networks of enterprises. This enterprise

knowledge can be used to identify new business opportunities and propose potential collaborations among enterprises. Bearing the above requirements in mind, an intelligent support system is proposed in this chapter to enhance and support the networked enterprises in successful, timely creation of and participation in a collaborative environment by alerting them with appropriate tendering opportunities and listing suitable partners for the possible projects.

To accomplish these functionalities, a platform is proposed in this chapter for the creation stage of a virtual enterprise. The main objective of this platform is:

To develop a relatively stable network of enterprises by raising awareness, and building and strengthening identity, trust and knowledge sharing among the participants.

The proposed platform uses the semantically enriched information about enterprises to raise awareness about business opportunities and enhance synergy among enterprises. The ECOS ontology introduced in the identification stage for capturing published competences of enterprises is used in this platform for analyzing and categorizing business opportunities. The platform identifies direct and indirect relationships among enterprises, based on their profiles and requirements of identified business opportunities. Enterprises can utilize these relations as a basis for building social relationships and developing a virtual community of enterprises willing to enter the global market in collaboration with each other. The platform is structured in three main modules i.e.

1. *ECOS-web*: This module captures, organizes and publishes enterprise competences in an explicit and machine readable format. The ECOS ontology proposed in chapter 4 is used to provide an explicit semantic description to all information resources and encode them in an unambiguous and machine understandable format. To facilitate the analysis of ECOS profiles, a network analysis and visualization toolkit is introduced in this module that gives a visual and mathematical analysis of enterprise relations,

ranging from general links to unidentifiable relationships. The key features of this toolkit are:

- To visualize communication and other relationships between ECOS profiles by means of diagrams,

- To study the factors which influence relationships and to study the correlations between relationships,

- To draw out implications of relational data,

- To make recommendations for improving communication and workflow within and between organizations.

2. *ECOS-track*: The introduction of the internet in the business environment has allowed companies to realize their selling and buying activities as well as their cross company communication through the World Wide Web. With the support of the internet, different organizations e.g. companies, government agencies, academic institutions and service providers etc publish their tender needs electronically. The task of the ECOS-track module is to extract the requirements (e.g. location, project requirement, date of submission, etc) from tender documents and weigh them against the interests and competences of enterprises, published as ECOS profiles.

3. *ECOS-match*: The ECOS-track module lists the potential enterprises that can fulfil the requirements of a tender. However, there can be several companies that might play a part in a project by collaborating with other enterprises, but could not undertake the tender requirements individually. The ECOS-match module categorises ECOS profiles based on the project requirements and sketches potential collaborations between enterprises. It provides a list of potential virtual enterprises, with a single enterprise or multiple members, fulfilling the needs of the tender.

The rest of this chapter is organized as follows: Section 2 to Section 4 provide details of each module respectively. The relevant literature associated with each module is provided separately in each section. A summary of this chapter is provided in Section 5.

5.2. ECOS-web: Maintaining Collaboration Pool

To explore opportunities in a global market, enterprises publish their skill, capability, expertise and competence that they are able to offer. The ECOS-web module registers the published competences of companies and thereby creates a pool of enterprises that are willing to collaborate and are potentially available and suitable for the creation of a virtual enterprise. In this research, the term *collaboration pool* is used to refer to a group of individually collaborating companies, individuals, academic organizations, customers, etc (Swarnkar, et al. 2009).

5.2.1. *Background and Motivation*

Enterprises, in general, publish their competences in the market or on the web to extend their business impact and develop collaborations with similarly focused companies. The Web has emerged as a basic entity for interconnecting man and machine. It has enabled firms to make their information available electronically so that users can access and share their resources and expertise and utilize a full range of data, knowledge and information. Despite growing interest and efforts, the current web based technology is still primitive in its functionalities. The next generation of web technology needs sets of interconnected data and semantic models to access heterogeneous and independent data repositories automatically by using scripts or programs. The power of the web has improved in two different directions: (1) as a *social web* connecting people and (2) as a *semantic web* connecting knowledge, as shown in Figure 5-1.

These two developments are following two different approaches for the same goal i.e. *to develop the ubiquitous web that will facilitate knowledge sharing, reuse, interoperability and collaboration*. The Semantic web uses defined meanings and formal ontologies for interconnecting computers and people. On the other hand, the Social web,
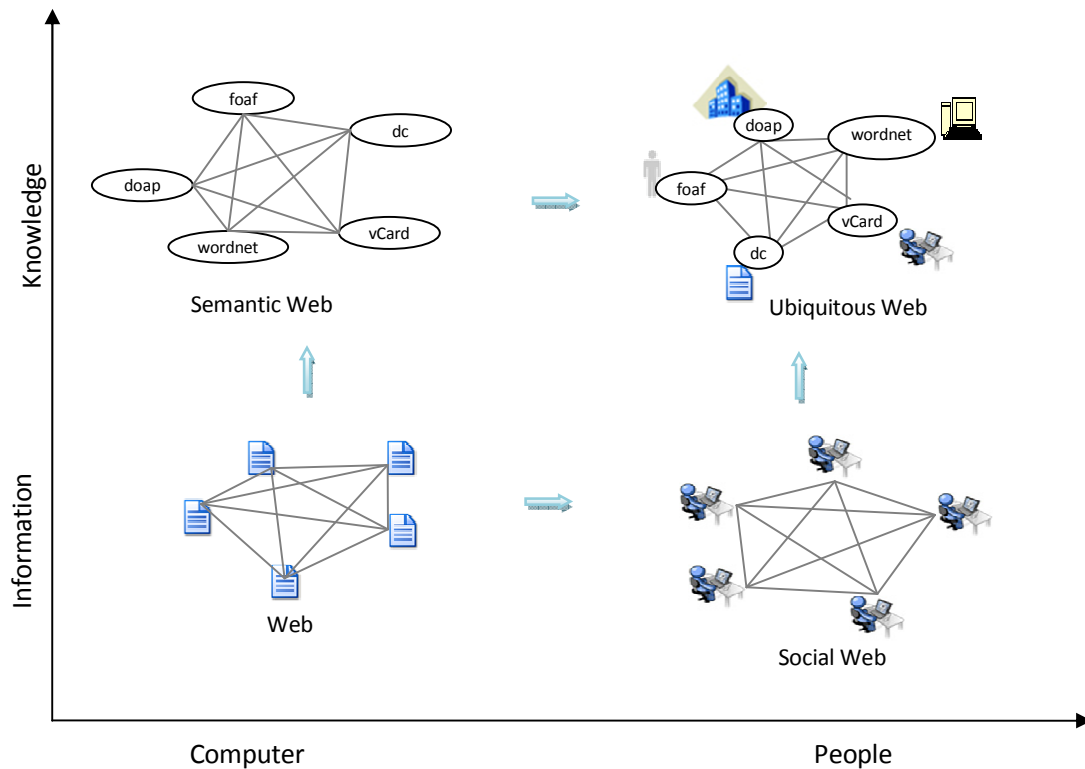
Figure 5-1: The Development of Ubiquitous Web from two approaches

part of Web 2.0, is mapping and measuring familiarities, relationships and flows for connecting people, groups and organizations. These two approaches are complementary and each has a lot to offer to the other. On one hand the social web has a wealth of interconnected data but lacks structured semantics and quality control. In contrast, the semantic web handles semantics, but suffers a dearth of users' interest. There is a need to connect these two approaches and develop a ubiquitous web connecting intelligence.

This research demonstrates such an alignment by the development of the *ECOS-web*, which is a step towards developing an enterprise social networking community on the semantic web. The scope of *ECOS-web* is the enterprise community willing to enter the global market and wanting to publish skills, capabilities, expertise and competence that they are able to offer. Before getting into details, a brief description of social networking,

enterprise social networking and a literature survey on the integration of the semantic web and social networking is provided.

5.2.1.1. *Social Network Analysis* (SNA)

Social network analysis is a field of research focusing on relationships among social entities such as members of a group, between corporations, or between nations. It is a formalism, primarily based on graph theory (Harary 1969), that provides a visual and mathematical analysis of interdependencies between information or knowledge entities with nodes and links. The nodes in the network represent actors (mostly knowledge entities such as people, organization etc), while the links show relationships or flows between the nodes. It is used to visualize and conduct mathematical analysis on the network. A visual representation of a network provides a rich and ecological understanding of a knowledge domain. Mathematical analysis is performed to identify interaction patterns among the network members, the number and structure of the subgroups within the networks and their evolution (Anklam 2003).

5.2.1.2. *Organizational Network Analysis* (ONA)

Organizational network analysis, an SNA discipline, is used for studying formal and informal networks within or between formal organizations. It is a quantitative and descriptive technique for creating statistical and graphical models of the people, tasks, groups, knowledge and resources of organizational systems (Merrill, et al. 2007). ONA is applied for assistance in process improvement, enterprise collaboration, market study, workforce diversity analysis, reengineering, etc.

There has been an increased interest in this methodology to map and analyze the nature and role of formal and informal relationships in enterprise networks. Using organization network analysis, Cross and Prusak identified four role-players that contribute to the effectiveness of an organizational network and play a critical role in its productivity (Cross, Prusak 2002). These key players include: central connector, boundary

spanner, information broker and peripheral specialist. Krackhardt and Hanson showed three types of relationship networks: advice network, trust network and communication network (Krackhardt, Hanson 1993). Similarly, Karen Stephenson, a social network theorist, identified six core layers of knowledge each having its own informal network (Stephenson 2006). These six layers include: work network, social network, innovation network, expert knowledge network, carrier guidance/ strategic network and learning network. Pedersen introduced the social network perspective for the analysis and management of a virtual organization (Pedersen 2007). In particular, Pedersen examined how the social structure of inter-organizational relationships may enable or constrain the formation and cooperation in virtual organizations.

5.2.1.3. *Social Network and Semantic Web*

In practical terms, social networks are web based social communities such as Facebook, MySpace, Orkut, LinkedIn etc (Shen, et al. 2006, Kim, et al. 2010). Social networking is a platform on the web where a group of people launch a highly interactive service based on common interests between users and easy to use communications tools to detail and promote those interests to others (Bloch, 1999). This new web technology has created a new generation of web based communities with a variety of ways to network, such as e-mail and instant messaging services. The inherent features of the web to publish and gather information are a major factor in the success of networking sites on the web. Web based social networks have grown quickly in number and scope since the mid 1990s (Golbeck, Rothstein 2008). There are about 250 websites dedicated to social networking providing explicit support for users to build, browse and interact with others (Hutchison 2008).

Even though social networking has grown dramatically on the web, these services are still primitive in their functionalities. Current web technology provides no information about the semantics of described content (Casey, Pahl 2003). It is based on an individual programmer or interpreter understanding the semantics of the language from available

descriptions. Such lack in information control and semantic description in meaning of contents in a particular domain has hampered the interoperability and sharability of knowledge (Lin, Harding 2007). These problems have been addressed in literature using semantic web technology. The Semantic web is defined as the conceptual structuring of the semantics of data in a machine readable way that enables web entities to interoperate with each other, dynamically discover resources, extract knowledge and solve problems (Guha, McCool 2003).

Even though the importance of the Semantic Web has been recognized and widely accepted by industry and academic research, a significant amount of research is still required to achieve a semantic environment in the social networking community. Considering the present technological scenario and bearing market requirements in mind, this research aims to develop a semantic web toolkit, ECOS-web, for registering enterprise competences and visualizing and analyzing them in the context of collaboration with other companies. This semantic toolkit will allow companies to post their published competences directly into online networked information spaces and become a part of a virtual community.

5.2.2. *ECOS-web Definition*

To fully exploit the skills and capabilities of an enterprise in a market, it is necessary to consolidate and contextualize enterprise competences in an explicit and structured manner (Sanchez 2004). The ECOS-web module identifies relations among enterprise profiles and provides a visual and mathematical analysis of these relations. This module utilizes the ECOS ontology to conceptually structure the semantics of enterprise profiles in a machine readable way and identifies relations among them. As discussed in Chapter 4, the ECOS ontology provides a set of complementary concepts and a comprehensive vocabulary for representing enterprise competences as a semantic web resource. This ontology is used as a metadata for creating machine readable profiles of companies and utilizing them in the ECOS-web module and other two modules which will be described in

Sections 5.3 and 5.4. The ECOS-web module maps and measures these ECOS profiles and provides a mathematical and visual analysis of enterprise relationships.

Conceptually, ECOS-web is defined as a network with a set of nodes representing ecos concepts and a set of edges representing one or more relations among them. Formally ECOS-web is defined as $W = (G, O, L)$, where G is a graph representing a collaboration pool (e.g. SMEs, firms, institutions, academics, etc), O is supporting ontologies for standard classifications, and L is collection of link sets which generalize the idea of a link, $L = (G \rightarrow O)$, between enterprise and supporting ontologies. An enterprise graph is defined as $G = (V, E)$ where V is a finite set of vertices representing enterprises as nodes in the collaboration pool and E is the finite set of edges that are constituted by node pairs $e = (u, v)$. The cardinality of V and E is denoted by n and m, respectively.

5.2.3. *ECOS-web Architecture*

In technical terms, ECOS-web is a network analysis and visualization toolkit. It consists of 4 layers, named ASAP i.e. Acquisition, Storage, Analysis and Presentation. The detailed explanation of ASAP is as follows:

1. *Acquisition*: This layer is concerned with data acquisition. ECOS-form, a web application mentioned in Chapter 4, is used for capturing enterprise competences and presenting it in RDF/XML format.

2. *Storage*: The RDF/XML format, used for representing an ECOS profile, is the most common approach for linking data with ontologies and its benefits are explained in section 4.3.4.

3. *Analysis*: The next step involves matching of semantic profiles captured as instances of the ECOS ontology and identifying relationships among them. Given a collaboration pool $EP = \{E_1, E_2, \ldots, E_P\}$, for any two enterprises E_i and E_j there could be a myriad of relationships linking the two. The idea is to identify significant links between any two enterprises in the collaboration pool and create an ECOS-web

network graph. Here, the significant link signifies relations that may affect enterprise relationships within the enterprise pool. The two categories of links considered in this module are:

- *Direct link* which refers to the direct connectivity of E_i and E_j. The semantics of the ECOS ontology includes certain properties that can direct an enterprise toward the ECOS profile of other enterprises. For example, *ecos:Customer*, *ecos:Relation* and *ecos:PastProject* include relations with profiles of other enterprises. Direct links are used to represent these straight relations.

- *Hidden link* is characterised by the similarity of a specific notion. As mentioned earlier, ECOS ontology labels enterprise information and competences with predefined classification codes. These classification codes are an exhaustive and structured set of mutually exclusive and well described information, generally presented as a hierarchy of numerical or alphabetical codes. The ECOS properties using these codes are linked with an unseen relation via a classification tree, presented with hidden links. These hidden links are linked with weight $w = 1/min\left(dist\left(code_i, code_j\right)\right)$, where $code_i \rightarrow e_i$ and $code_j \rightarrow e_j$, i.e. reciprocal of distance between two codes in a classification hierarchy.

The above analysis results in a network graph where nodes are enterprises, links are the relations between the nodes and weights are the strength of the relations. For example, a network graph with 6 enterprise nodes and links for customer relations is shown in Figure 5-2. In this ECOS-web graph, the yellow nodes are the company names, the blue nodes are for customers and their enterprise codes and blank solid lines are used to link enterprises with their customers (e.g. ABCJ is customer of BCRA and Anchor). Similarly, other graphs can be created for various other direct and indirect links, such as project partners, similar processes and products, etc. In the ECOS-web module, these graphs are combined together to get a single weighted network graph. The primary use of weighted graph in the ECOS-web
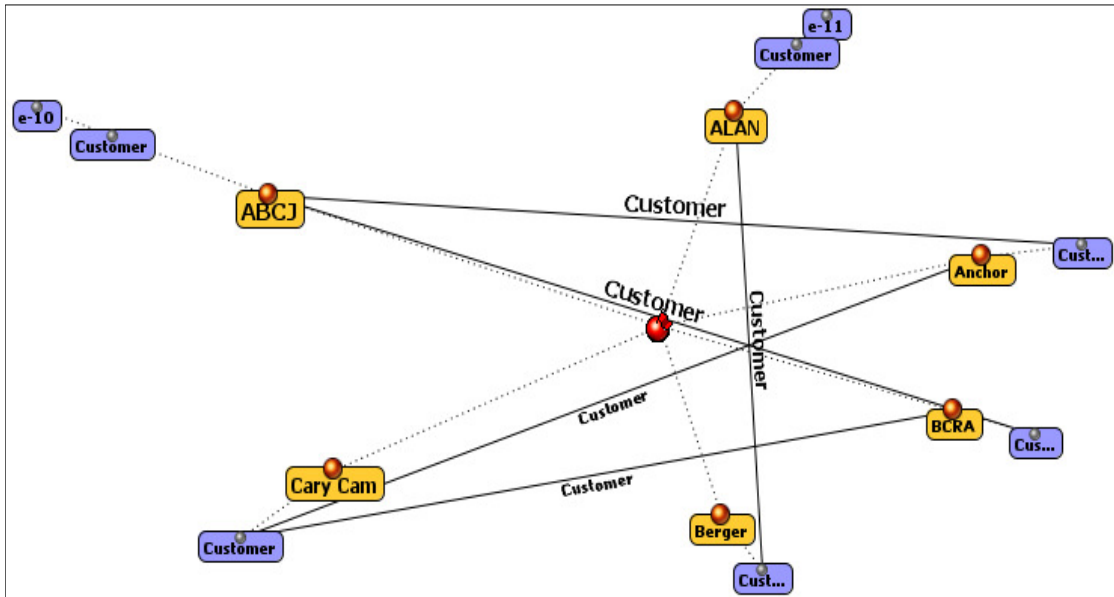
Figure 5-2: A network graph with links for customers

is to identify the importance of an enterprise and their pattern of relationships in a network. The analysis of ECOS-web graph can assist users in the identification of enterprises which are more important in the network. It can be used to identify enterprises which have more ties with others, can reach all the others more quickly, have control the sharing between the others, tend to group together, etc.

The major concept used for analyzing the weighted network graphs is centrality of the graph. Centrality is useful for understanding power, stratification, ranking and inequality within a network graph. It is the measure of how close a node is to the centre of action in the network. Freeman formalized three different measures for measuring node centrality that are degree, closeness and betweenness (Freeman 1979). However, the measures defined by (Freeman 1979) are only designed for binary networks in which links are either present or absent. There have been a number of attempts to generalize these centrality measures for weighted graphs (Barrat, et al. 2004, Brandes 2001, Newman 2001). These attempts mainly focused on link weights and not on the number of links, which formed the basis of the original

measures in (Freeman 1979). Opsahl et al proposed generalizations that combine both the aspects, i.e. number of links and weights assigned in calculating the centrality measures (Opsahl, et al. 2010). For a weighted network graph $G = (V, E, w)$ with *n* vertices, these modified measures are calculated as:

- *Degree*: Measuring the degrees of the nodes in the network is the simplest and most straight forward way to determine centrality. It is a basic indicator that determines the number of nodes a node is connected to. Freeman defined this measure for node *v* in graph *G* as (Freeman 1979):

$$C_D = {\sum_{u}^{n} x_{uv}}\big/{n-1} \qquad\qquad \text{5-1}$$

where, x_{uv} is the number of links incident upon node v , i.e. $x_{uv} = 1$ if the node v is connected to u or 0 otherwise. (Barrat, et al. 2004, Brandes 2001) modified this measure for weighted graphs and formalized the measure as follows:

$$C_D^w = {\sum_{u}^{n} w_{uv}}\big/{n-1} \qquad\qquad \text{5-2}$$

where, w_{uv} is the weight assigned to the link between node v and u. Opsahl et al. combined these two formalizations by adding a tuning parameter to determine the relative importance of the number of links compared to link weights (Opsahl, et al. 2010). The modified measure is given as:

$$C_D^{w\alpha} = C_D^{1-\alpha} \times C_D^{w\,\alpha} \qquad\qquad \text{5-3}$$

where, α is the tuning parameter defined set as per the research context and data by the users. In the ECOS-web, it is used to finding enterprises with the most direct connections to others. An enterprise with high degree is:

- Connector or hub in the network

- Can be identified as third parties or deal makers

- *Closeness*: Focuses on how close an actor is to all other actors in the network. Freeman defined it as the mean geodesic distance between a node v and all other nodes in the network (Freeman 1979):

$$C_C = \frac{1}{\sum_{t \in V \backslash v} d(v,t)}$$ 5-4

where, $d(v,t) = min(x_{vi} + \cdots + x_{it})$ is the minimum distance between node v and t with h as the intermediary node. Barat et al. modified this distance measure by replacing the binary values for links with their weights (Freeman 1979, Barrat, et al. 2004), given as:

$$d^w(v,t) = min\left(\frac{1}{w_{vi}} + \cdots + \frac{1}{w_{it}}\right)$$ 5-5

Opsahl et al further modified this weighted distance measure by including the tuning parameter (Opsahl, et al. 2010), given as:

$$d^{w\alpha}(v,t) = min\left(\frac{1}{w_{vi}{}^\alpha} + \cdots + \frac{1}{w_{it}{}^\alpha}\right)$$ 5-6

Finally, the closeness centrality measure with modified shortest path function is given as:

$$C_C^{w\alpha}(v) = \frac{1}{\sum_{t \in V \backslash v} d^{w\alpha}(v,t)}$$ 5-7

In an ECOS-web graph, an enterprise with a high closeness centrality generally:

- has quick access to other entities in a network.

- is close to other entities.

- has high visibility as to what is happening in the network.

- *Betweenness*: Is the extent to which a particular enterprise node lies between the various other nodes in the network. For a graph G, the betweenness for node v defined by (Freeman 1979) is:

$$C_B(v) = \sum_{\substack{s \neq v \neq t \in V \\ s \neq t}} \frac{\sigma_{st}(v)}{\sigma_{st}} \qquad\qquad \text{5-8}$$

where, σ_{st} is the number of shortest paths from s to t and $\sigma_{st}(v)$ is the number of shortest path from s to t that pass through v. Opsahl et al modified this measure by using the modified distance measure (equation 5-6) (Opsahl, et al. 2010), given as

$$C_B{}^{w\alpha}(v) = \sum_{\substack{s \neq v \neq t \in V \\ s \neq t}} \frac{\sigma_{st}^{w\alpha}(v)}{\sigma_{st}^{w\alpha}} \qquad\qquad \text{5-9}$$

An enterprise with a high betweenness centrality generally:

- has ability to make connections with other enterprises in the network.

- holds a favoured or powerful position in the network.

- has a great amount of influence over what happens in the network.

In this research the modified measures are used for calculating the centrality of ECOS-web graphs.

4. *Presentation*: The presentation layer of the ECOS-web is for browsing and visualizing the network graph. Visualization tools have always been key elements in network analysis. They allow users to see data with new insight that is unsurpassed by summary statics. Network analysis has made extensive use of visualization since Moreno first introduced the sociogram (Moreno 1937). A sociogram is a representation of network connections of two or more with entities. There are a number of variations on the theme of sociograms, but they all share the common feature of using a labelled circle for each entity in the network and line segments between entities to represent relations that exist between the two. In this module, a sociogram is used for visualizing the ECOS-web and analyzing enterprise relation graphs.

The real life implementation of the ECOS-web module is shown in Chapter 8.3.

5.3. ECOS-track: Prioritizing Business Opportunities

The main task of the ECOS-track module is to create awareness about new business opportunities among enterprises. In order to better understand the role of ECOS-track in virtual collaborations, consider the following example scenario:

> A company, XYZ, is in the business of supplying automotive aftermarket for electrical consumables. Its core competence lies in its customer focused staff that understand current and future market requirements and provide customers with an increasing competitive advantage in the ever changing automotive supply market. Being an SME with 25 full time employees, the company always seeks new projects and appropriate requests for tenders appearing in the market. In order to ease their search, the company has registered with different tender-alert services e.g. Tender Electronic Daily, Sell2Wales etc. Each morning, one of the managers of the company receives emails with lists of suitable tender opportunities for his firm. Then his job is to manually scan each tender opportunity and sort them according to their requirements and suitability for the current situation of his company. A list of possible opportunities is then passed to a higher level for discussion and selection in his company.

It is evident from this scenario that online tender scanning and selection of suitable tenders is of significant importance for enterprises. Large companies, generally, employ groups of people from technical, financial, service–related fields to evaluate tender proposals for feasibility. However, small enterprises cannot always afford to spend too much of their resources on evaluating tender opportunities. At the same time, in order to maintain their reputation and to remain in business, they cannot afford to lose opportunities. For an enterprise with only a handful of personnel, certain officials are likely to have to evaluate tenders in addition to their other duties. These members of staff can easily spend 0.5-2 hrs daily looking for suitable tender opportunities that are available in market, and it is quite difficult for them to manage this time requirement along with other responsibilities (Swarnkar, et al. 2009).

Companies may register with tender alert services which identify and list possible requests for tenders based on the resemblance of tenders with information provided by the company. It is therefore an important task for a subscriber to publish an appropriate set of competences and preferences of their company. Obviously, an enterprise would like to receive notifications about all relevant tenders in their area of core expertise.

Additionally, they may like to receive alerts for the tenders that are closely related to their core expertise. However, this increases the number of sectors they can get alerts from and hence escalates the number of alerts received every day. With a large number of notifications to work through, the process of tender evaluation becomes an elimination process rather than a selection process.

Considering the above requirements the ECOS-track module has been developed for prioritizing requests for tenders. The main functionality of the proposed module is four-fold:

1. Capture enterprise profile information, such as who it is, what it can do, what it is interested in, what it wants to achieve etc.
2. Track new business opportunities and subsequently extract the relevant information (e.g. contract authorities, scope and duration of tender, etc) from the tender documents.
3. Weight tender documents based on their requirements and the interests and competences of enterprise.
4. Notify the enterprise with a prioritized list of tenders.

Before getting into a detailed description of the ECOS-track module, a brief review of the published work on tender selection problem is provided in next subsection.

5.3.1. *Related Literature*

A number of different approaches can be found in literature for tender selection and prioritization. According to Weber et al., the selection of tender and vendor greatly affect the success of an organisation in the market and is one of the most important purchasing decisions (Weber, et al. 1991). Weber et al. gave a review of 74 examples of research work concerned with the vendor selection process. The research articles in (Weber, et al. 1991) are mainly grouped in three main categories, namely, linker weighting models,

mathematical programming models and statistical/probabilistic approaches. Additionally, a variety of different approaches have also been introduced in recent years.

Wong et al. investigated a multi–criteria selection approach as opposed to the lowest price criteria for the construction industry (Wong, et al. 2001). Ghodsypour and O'Brien presented a mixed integer non-linear programming model to solve the multiple sourcing problems, taking into account the total cost of logistics including net price, storage, transportation, and ordering costs (Ghodsypour, O'Brien 2001). An approach for solving qualitative multi–criteria analysis by fuzzy pair-wise comparison is presented by Deng (Deng 1999). Hsieh et al. present a fuzzy multi-criteria analysis approach for selecting planning and design alternatives in a public office building (Hsieh, et al. 2004). Wang et al. applied incomplete linguistic preference relations to evaluate tender selection criteria and provide decision matrices for making pair-wise comparisons (Wang, et al. 2007).

The research work listed above addresses the problem of tender selection from the point of view of the contracting authority, e.g. local governments or larger companies outsourcing its projects. In contrast, only a few examples of research work approaches to the tendering problem from the bidder's perspective. Swarnkar et al proposed a knowledge discovery and mining aided multi–criteria decision making approach to assist the tender offer selection process (Swarnkar, et al. 2009). Similarly, the ECOS-track module is proposed in this research to facilitate the tender selection process by prioritizing the tenders for companies.

5.3.2. *ECOS-track Process*

The step wise description of the approach followed in the ECOS-track module is as follows:

1. At first, the skills and capabilities of an enterprise required for tender selection is identified and stored. In this research ECOS-card, a semantic profile based on ECOS ontology, is used for capturing the competences of the enterprises. The ECOS

Table 5-1: Ratings for CPV Comparison

Rating	Tender CPV	Enterprise CPV	Explanation
1	XXXXX YYY-Z	XXXXX YYY-Z	Complete code is similar
0.75	XXXXX YYY-Z	XXXXX 000-0	First 5 digits are similar
0.5	XXXXX YYY-Z	XXXX0 000-0	First 4 digits are similar
0.25	XXXXX YYY-Z	XXX00 000-0	First 3 digits are similar
0.0	XXXXX YYY-Z	XX000 000-0	First 2 digits are similar

ontology includes a wide range of factors defining skills and capabilities of an enterprise. However, the ECOS-track module uses a limited set of factors for tender selection that include a company's items of interest such as preference for location, region, products or process.

The relevant information taken from the ECOS-card is matched with the requirements of the available business opportunities. A business opportunity is generally published as a tender document, available as html page, PDF file, word document, etc. Such e-documents include a wide range of information, e.g. basis of the tender, background to the work, user experience, peripherals and ancillaries, deliverables, format / specification, validity period, etc. However, only a limited set of information is taken for tender selection, i.e. contracting authorities, object of contract, required CPV codes (Common Procurement Vocabulary), its scope and duration, location etc.

2. Given an enterprise E with a set of preferences $E = \{P_1, P_2, \ldots, P_N\}$, and a set of tenders $T = \{T_1, T_2, \ldots, T_L\}$ with each tender having a set of selection criteria $T_l = \{C_1, C_2, \ldots, C_M\}$. Each individual requirement of a tender $C_m \in T_l$ corresponds to a certain selection criteria and is weighted against a suitable set of preference, e.g. location is evaluated against location but not against duration or CPV codes. The individual rating R_{lm} for a tender requirement C_m of tender T_l is evaluated as:

$$R_{lm} = \forall_{n=1}^{N} max(x_{mn} \times r_{lmn}) \qquad\qquad 5\text{-}10$$

where, $l = 1, \dots, L$; $m = 1, \dots, M$; r_{lmn} is the weight assigned to a criteria C_m of tender T_l against the list of preferences;

$$x_{mn} = \begin{cases} 1; \; if \; C_m \; can \; be \; compared \; with \; P_n \\ 0; otherwise \end{cases}$$

The numerical value of r_{lmn} is calculated differently, depending on the type of selection criteria. For example, the sector is compared with the Boolean logic i.e. $if \; C_m = P_n \Rightarrow r_{lmn} = 1; else \; 0$; the location is compared by calculating the distance between two NUTS code in NUTS tree classification; the CPV codes are compared as per the ratings defined in Table 5-1. In this research, the CPV codes are the key information used for tender matching (**Appendix B**). This 8 digit codes with a standard vocabulary is used to facilitate fast and accurate comparison of tender notices against enterprise competences.

Then the total utility of a tender document is calculated, given by

$$U_l = (\textstyle\sum_{m=1}^{M}(R_{lm} \times w_m))/M \qquad\qquad 5\text{-}11$$

where, w_m is the weight assigned to a selection criteria as per its importance for an enterprise; $0 \leq U_l \leq 1$.

3. Finally, the enterprise is notified and sent a list of tender documents prioritized as per their utility ratings.

It is apparent that an enterprise can apply for a tender T_l individually, if its utility factor $U_l \approx 1$; otherwise enterprise has to create a virtual collaboration to accomplish the tender requirements. The ECOS-match module analyzes competences of other enterprises and suggests potential collaborations that can fulfil the tender requirements completely.

5.4. ECOS-match: Suggesting Enterprise Collaborations

The selection of appropriate business partners and sharing of resources with them is a major challenge associated with the creation of a virtual enterprise. There are many factors that affect the partner selection process such as identity, trust, knowledge sharing, etc. However, the ECOS-match module mainly focuses on partner selection based on the requirements of a tender, which are not satisfied by an enterprise and for which they need partners. The objective is to form a virtual collaboration of appropriate partners coming together to fulfil all the requirements of the tender. It is apparent that there can be many possible combinations of partners satisfying the requirements of a tender, thus the links (direct and hidden) identified in the ECOS-web modules are used for selecting the best and appropriate combination of partners. Consequently, the enterprises with strong direct relationships and fewest hidden relation (i.e. less similarity of competences) will have greater chance of working together.

5.4.1. *Partner Selection in Virtual Enterprise*

The Partner selection process can be described as follows.

> Assume that an enterprise XYZ is planning to tender for a project. While going through the request tender documentation, the company official finds that it can complete a large part of the tender, but is not able to complete all the tasks asked by the tender. Therefore it needs help from other enterprises which can complete the required tasks in collaboration with itself by forming a virtual collaboration. The main job is to select the enterprises from the collaboration pool which can form a potential virtual enterprise for the tender.

The ECOS-track module matches a set of requirements from a tender against the profiles of enterprises with their set of products, skills, processes etc. It evaluates a utility factor based on the extent that an enterprise satisfies the request for tender. The partner selection problem can also be taken as the same problem, but there is a difference as in this case there are requirements of a tender which are not fully satisfied by the enterprise. Therefore, the competences of enterprises in the collaboration pool are checked against the remaining demands of the tender and then on the basis of their utility factor, appropriate partners are selected from a vast collection of enterprises. To

further increase the efficiency of the process, the similarity of the enterprise with others is also taken into account to remove the enterprises which are the direct competitors of the enterprise. The factors that have been taken into account during the ECOS-match process include the demands of the tender that are not satisfied by the enterprise, similarity of the enterprise, evaluated in ECOS-web module, with other enterprise and past relations, records and projects of the enterprise with other enterprises in the collaboration pool.

5.4.2.   *Related Work*

A brief literature review of some work done in the field of partner selection in virtual enterprise is provided in this section. Talluri and Baker proposed a two-phase quantitative framework to facilitate the decision making process of selecting an efficient and compatible set of partners (Talluri, Baker 1996). The first phase of the process identifies the efficient candidates for each type of business process (e.g. design, manufacturing, distribution, etc) utilizing data envelopment analysis. The second phase involves the execution of an integer goal programming model to determine the best portfolio of efficient partners based on a number of compatibility objectives. Feng et al. and Cao and Weng proposed a multi-objective optimization model with cost, due date and precedence of sub-projects as basic factors for partner selection (Feng, et al. 2000, Cao, Wang 2002). Fuqing et al. (Fuqing, et al. 2006) modified the model proposed in (Feng, et al. 2000) and (Cao, Wang 2002) by adding a factor for risk of failure of the project and used R-GA with embedded project scheduling for solving the partner selection problem.

   Hajidimitriou and Georgiou presented a quantitative model, based on the goal programming technique, to evaluate potential candidates and select partners for forming a joint venture (Hajidimitriou, Georgiou 2002). The criteria used in this model are ability of the partner to gain entry into local market; political advantage to secure access to raw material resources; partners' familiarity with local business practices, politics and

customs; ability to facilitate exporting, ownership of patents, licenses etc. Wang and Xiong converted the partner selection problem into an integer programming problem with non-analytical objective function (Wang, Xiong 2001, Wang, et al. 2002). They proposed a fuzzy embedded genetic algorithm based approach to solve the partner selection problem and compare the results other existing algorithms.

5.4.3. *ECOS-match Process*

The detailed description of the approach followed in ECOS-match module is provided in this section. As mentioned earlier, the ECOS-match does not just search the group of enterprises that can fulfil all the requirements of a tender; but also considers the relations identified in ECOS-web module. Therefore, the partner selection is solved considering three basic aspects i.e. utility factor of tender, direct and indirect relations among enterprises. A step-wise description of the approach is as follows:

1. Given a collaboration pool $EP = \{E_1, E_2, \dots, E_P\}$ with P enterprises willing to participate in virtual collaborations. The pool is developed by capturing the published competences of enterprises in ECOS format. Let $w_{(dl)}$ and $w_{(hl)}$ be the weights assigned for the direct relations and hidden relations among the enterprises, respectively.

2. Let $T_l = \{C_1, C_2, \dots, C_M\}$ be the tender document with M requirements. The utility factor for T_l evaluated against E_p is represented as $U_l(E_p)$. The utility factor must be $U_l(E_p) < 1$, such that $1 - U_l(E_p)$ is sufficiently large for E_p to go for virtual collaboration and look for partners with complementary competences.

3. The next step is to analyze all combinations of pool members that can collaborate and form a virtual enterprise. Brute-force search is used to try all possible 2^p combinations ($2^p = C_0^p + C_1^p + \dots + C_p^p$) of enterprises. Brute-force search is a trivial but very general problem-solving technique that systematically enumerates all possible candidates for the solution and checks whether each candidate satisfies the

problem's statement. A multi-objective function is proposed in this research to identify the best possible virtual enterprise by using the above search method.

Let $VP = \{E_1, E_2, \ldots, E_V\}$ be a virtual enterprise with V partners such that $VP \subseteq EP$ and $1 < V \leq P$. The objective function used for evaluating VP is given as:

$$Obj_{VE} = (w_1 \times Net\ Utility) + (w_2 \times Net\ Direct) + (w_3 \times Net\ Hidden) \quad \text{5-12}$$

where, $Net\ Utility$ is the utility factor tender T_l calculated against virtual enterprise VP; $Net\ Direct$ is the average of weights for direct relations among VP members; $Net\ Hidden$ is the average of the weights for hidden links among VP members; w_1, w_2 and w_3 are the weight assigned to the respective factors, such that $-1 \leq \{w_1, w_2, w_3\} \leq 1$. The three objectives are calculated as follows:

$$Net\ Utility = \sum_{m=1}^{M} w_m \times \left(\left(\sum_{v=1}^{|VE|} R_{lm}(E_v)\right) / \left(\sum_{v=1}^{|VE|} y_{mv}\right)\right) \quad \text{5-13}$$

where, $R_{lm}(E_v)$ is the rating assigned to the tender requirement C_m for T_l against the acquirements of E_v;

$$y_{mv} = \begin{cases} 1; & if\ E_v\ satisfies\ C_m \\ 0; & otherwise \end{cases}$$

Here individual rating are utilized, rather than the utility factor U_l, to minimize the rating of the tender requirements fulfilled by more than one enterprise. The other two factors are:

$$Net\ Direct = \left(\sum_{v=1}^{|VE|} \sum_{u=1}^{|VE|} w_{uv(dl)}\right) / (|VE| \times |VE|) \quad \text{5-14}$$

$$Net\ Hidden = \left(\sum_{v=1}^{|VE|} \sum_{u=1}^{|VE|} w_{uv(hl)}\right) / (|VE| \times |VE|) \quad \text{5-15}$$

Finally, all the combinations of enterprises satisfying the requirements of a tender document are evaluated and sorted as per their objective values. The pool members in the selected combinations are notified about the new business opportunity and potential partners.

5.5. Summary

During the creation stage, a stable network of enterprises is developed to respond and make decisions related to the new business opportunity coming in the market. At this stage, a knowledge management framework is required to raise the awareness of issues that may benefit the creation of the virtual enterprise. In this chapter, a semantic web platform is developed to facilitate the creation of a virtual enterprise by raising awareness about the new business opportunities and suggesting potential partners for the collaboration. The platform is structured in three main modules: ECOS-web – to maintain the collaboration pool, ECOS-track – to track business opportunities and ECOS-match – to categorize the pool and create awareness in the pool. The three modules use the ECOS profiles introduced during the identification stage to identify relations with business opportunities and other enterprises' profiles.

The detailed implementation of the proposed platform on an experimental case study is shown in Chapter 8.

C h a p t e r   6

OPERATION STAGE: SEMANTIC ANNOTATION OF DOCUMENTS

Information is the seed for an idea, and only grows when it's watered
Heinz V. Bergen

ABSTRACT

The pervasive challenge during the operation stage of a virtual enterprise is to provide correct and timely information to the right people at the right time. However knowledge sharing and communication in a collaborative environment is complex and challenging, as enterprises often use domain specific keywords and phrases to describe their information resources. A specialized knowledge service, TEXT2RDF is proposed in this chapter that gleans information from documents, hyperlinks the terms and keywords with domain specific meanings and converts in an unambiguous and machine understandable format.

CHAPTER OUTLINE

6.1. Introduction

> In an agricultural economy land is the key resource. In an industrial economy natural resources, such as coal and iron ore, and labour are the main resources. A knowledge economy is one in which knowledge is the key resource. (Houghton, Sheehan 2000)

The current economy, referred to as a service economy, digital economy, or network economy, is emerging as a knowledge and idea-based economy with knowledge and intangible resources as valuable organizational assets (Robinson 2004). The difference with this emerging economy is the magnitude of information and knowledge incorporated into economic activities that has changed the basis of competitive advantage (Fragidis, et al. 2008). In the past, syndication was less of a challenge because in more local markets everybody was aware of what others do. However in the current business climate, when the market is big and companies are realising the *virtual enterprise* vision, knowledge is distributed throughout, rather than being centralized. Distributed knowledge means that no one is capable of achieving the highest level goal(s) on his own; goal achievement becomes a team activity, where different enterprises share information for the achievement of goal(s).

Compared to single enterprises, information sharing and communication in a virtual enterprise is a notoriously challenging and complex problem (Linthicum 1999). In a virtual enterprise, although companies perform the operations using their core competencies, they also share and exchange relevant information with other collaborators. However, collaborating partners often prepare and store documents, with different terminologies describing the same meaning or the same terminology is associated with different meanings. The use of domain specific terminologies and phrases for describing the meaning of contents hampers the interoperability and sharability of knowledge between collaborative firms. A specialized knowledge service is required that can annotate domain specific knowledge to create machine readable information resources and thus reduce the current risk of misinterpretation.

The idea is:

To provide explicit semantic descriptions to all information resources and annotate them with domain specific keywords and phrases.

The traditional approaches for annotating the web resources with machine processable metadata require human interference and thus make the annotation process time consuming and inconsistent. Although, the semantics of web documents are difficult to decide without the judgement of human intervention, the automation of a semantic annotation process is necessary since even minor requirements of human effort will result in tremendous amounts of time and cost. However, such automation is rather difficult because the recognition of semantics is a high-level cognitive process (Yang 2009). According to (Yang 2009), a suitable text mining approach with the following requirements may provide a solution to this automatic semantics extraction problem:

- Text mining process should be fully automatic without or with a negligible amount of human intervention.
- The process should be generalizable and scalable so that it does not use existing web resources for training.
- It should extract the real semantics of the web documents and present it in a human comprehensible manner.

Bearing the above factors in mind, this research presents a semantic extraction method that searches for and extracts specific information from web documents and adds semantic metadata and tags to it. A text mining approach, TEXT2RDF, is proposed in this chapter to glean information from documents and converts it into a semantic web resource using RDF and RDF Schema. The terms and keywords extracted from the web documents are hyperlinked with domain specific information, available as a business directory, glossary, dictionary, etc.

The novelty of this research resides in providing an automatic knowledge extraction and representation approach, *for the first time*, on the semantic web. The proposed

concept of extracting terms and phrases from the web documents and hyperlinking them with domain specific knowledge is inspired by the idea of Linked Data, a sub-topic of the Semantic Web. The idea is, if enterprises publish their domain specific knowledge on the semantic web, the proposed TEXT2RDF approach can utilize this information for generating a semantic web document annotated with explicit enterprise knowledge. With such unified and unambiguous information resources, the virtual enterprise with partners from different backgrounds, culture or locations can easily be interoperated at the data level.

The rest of this chapter is organized as follows: Section 6.2 explains the need for semantic annotation in operation stage. Section 6.3 examines the requirements for managing information on the semantic web. In section 6.4, a detailed description of the proposed framework is provided. Section 6.5 explains the hyperlinking approach used for annotating the domain specific knowledge with the semantic web resource. Finally, the summary of this chapter is provided in Section 6.6.

6.2. Need for Semantic Annotation in Operation Stage

Once the process of selecting an appropriate business opportunity and suitable partners, and agreement on collaboration contract is finished in the creation phase (discussed in Chapter 5), the virtual enterprise enters the operation stage. The focus during the operation phase is on achieving the goals and objectives of the virtual enterprise. The partners come together to share information, resources, and capabilities and satisfy the functional requirements of the project. During this stage, partners are empowered with the effective capability of mutual communication and sharing of information, requests, results, proposals, etc. The sharing is often carried out over the computer networks and among autonomous, diverse and possibly geographically distributed enterprises.

The core requirement for efficient collaboration and sharing during the operation phase is to develop an interoperable system with seamless communication among distributed and heterogeneous computing entities representing different enterprises,

people and resources. In general, interoperability covers three domains (Berre, et al. 2004):

* *Technical interoperability* covers the technical issues of linking up computer systems and services. This includes key aspects such as open interfaces, interconnection services, data integration and middleware, data presentation and exchange, accessibility and security services.

* *Information interoperability* is used for resolving the semantic heterogeneity in the information systems. It is the ability for senders and receivers to interpret the information in the same way. This may require transformations of information representation or messaging sequences.

* *Organizational interoperability* captures the willingness of partners to undertake the necessary actions necessary for the collaboration. It is concerned with defining business goals, modelling business processes and bringing about the collaboration of administrations that wish to exchange information.

In this research the focus in mainly on information interoperability. The idea behind information interoperability is to allow computing entities to exchange information and understand its meaning, despite the fact that they were not originally conceived for cooperation.

During the operation phase of a virtual enterprise, the problem in ensuring information interoperability comes when partners with distributed and heterogeneous systems communicate and share information. The interoperability is often hindered because information is spread across many locations and is represented with varying degrees of structure using different formats and conflicting semantics. Different members of virtual enterprises usually use their own semantics classified in their local taxonomies, glossaries, dictionaries etc, which results in heterogeneous data, information, and knowledge representations. For example, the terminologies used by a manufacturing company will be different from the companies from other sectors, such as

construction, information technology, pharmaceutical, etc. Due to this heterogeneity, partners requesting the information have to understand the implicit semantics of the provider and choreograph them to establish interoperability, which evidently causes delay, unavoidable human errors and most importantly accuracy due to the lack of semantic processing (Muthaiyah, Kerschberg 2008).

In order to solve these problems, there is a need to identify the semantics of the information elements that are involved in the communication and attach local meaning to it. The TEXT2RDF application is proposed in this chapter to extract terms and phrases from the document and convert the web document into a semantic web document. Subsequently, the extracted terms and phrases are linked with the terminologies defined in the local classification systems of the enterprises, such as directories, catalogues and yellow pages etc.

Prior to getting into the details of the proposed TEXT2RDF approach and semantic annotation, the next section provides brief detail of the fundamentals of managing information on the semantic web.

6.3. Managing Information on the Semantic Web

It is commonly accepted that it is the ubiquity of web technology and the internet that has initiated and made possible the real time exchange of information on computers. The concept of the semantic web adds a new level to web technology (Dieng, et al. 1999). The semantic web is envisioned as an extension to the current web technology in which information is annexed with a well defined meaning to enhance the interoperability of computers and people (Casey, Pahl 2003). The declaration of domain knowledge in a machine readable way enhances the understandability of disparate information. The idea is to provide machine processable metadata that describes the semantics of resources and facilitates search, filtering, condensing, or negotiation of knowledge for human users.

6.3.1.   *Semantic Web: Building Blocks*

The vision of the Semantic web is to add meaning to the data so that computers can understand and perform intelligent search, knowledge representation and reasoning. The purpose behind adding meaning to a web resource is to make a computer understand the idea delineated in that resource. In order to understand the concept, the computer needs three basic things: *language*, *grammar* and *query language*. The basic building blocks needed for representing information on the semantic web are explained as follows:

1. *Semantic Web Language*: The Semantic web relies heavily on ontologies for defining semantics and giving meaning to the data and applications for automatic processing. Ontology is a form of knowledge representation comprising of a set of concepts, axioms, and relationships that describe a domain of interest. With the support of an ontology, a user can communicate with the system by developing and sharing a terminology for building knowledge bases for particular domains. Due to its strong implications in conception of reality, it has gained much interest in artificial intelligence for defining the basic terms and relationships using the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary (Lammari, Métais 2004). An ontology is developed and used for one or all of the following purposes:

    •   *Share knowledge*: sharing the structure of information among software agents and people,

    •   *Reuse knowledge*: reusing it for other systems operating in a similar domain,

    •   *Make assumption about a domain*: for easier communication and understanding.

    In this research, ontology is used for defining the semantics for web documents and utilizing them for explicit information representation on the semantic web. An ontology can enhance the manageability of large amounts of information resources, including web pages, text documents, email etc. The ontology model developed for
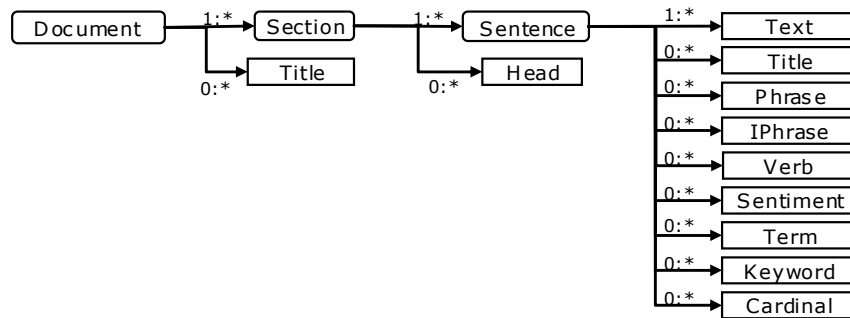
Figure 6-1: An Ontology Model for Document Representation

representing the web resources is shown in Figure 6-1. This model can help academic or industrial collaborators to capture, organize and share information in a unified and explicit manner, without any risk of misinterpretation.

2. *Semantic Web Grammar*: Informal ontologies hamper the effectiveness of interoperability and lead to ambiguities, delays and unnecessary work. Therefore, ontologies are encoded in a formal language for expressing the concepts in the domain, in a manner that computers can understand and manipulate easily. Several ontology languages have been developed in the past few years and used in the context of the semantic web, as discussed in Section 4.3.2. In this research, RDF is used for defining concepts and making simple relations between them.

RDF is a widely recognized standard for annotating online Web documents and for transforming the HTML Web to the Semantic Web (Davalcu, et al. 2003). RDF was specified by the WWW Consortium (W3C) and was originally designed as a metadata model so that it provides a syntax convention for representing the semantic of data in a standardized interoperable manner. The RDF metadata model is based upon the idea of making statements about Web resources in the form of "subject-predicate-object" expressions, called triples in RDF terminology. The subject denotes the resource, the predicate denotes property of the resource and object expresses a relationship between the subject and the object.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
         xmlns:base="http://www.yourURL.com/name/1/">
   <rdfs:Class rdf:ID="Document" />
   <rdfs:Class rdf:ID="Section">
       <rdfs:subClassOf rdf:resource="#Document"/>
   </rdfs:Class>
   <rdfs:Class rdf:ID="Sentence">
       <rdfs:subClassOf rdf:resource="#Section"/>
   </rdfs:Class>
   <rdf:Property rdf:ID="title">
       <rdfs:domain rdf:resource="#Document"/>
       <rdfs:range rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Literal"/>
   </rdf:Property>
   <rdf:Property rdf:ID="subject">
       <rdfs:domain rdf:resource="#Sentence"/>
       <rdfs:range rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Literal"/>
   </rdf:Property>
   <rdf:Property rdf:ID="subject">
       <rdfs:domain rdf:resource="#Sentence"/>
       <rdfs:range rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Literal"/>
   </rdf:Property>
   </rdf:Property><rdf:Property rdf:ID="verb">
       <rdfs:domain rdf:resource="#Sentence"/>
       <rdfs:range rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Literal"/>
   </rdf:Property>
   ...
   </rdf:Property><rdf:Property rdf:ID="cardinal">
       <rdfs:domain rdf:resource="#Sentence"/>
       <rdfs:range rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Literal"/>
   </rdf:Property>
   ...
</rdf:RDF>
```

Figure 6-2: RDF Schema for Ontology model shown in Figure 6-1

The RDF schema for the proposed ontology (Figure 6-1) is shown in Figure 6-2. This schema does not offer anything specific to the proposed framework, except for one key feature - a standard way to describe vocabularies and distribute them over the web.

3. SPARQL (SPARQL Protocol and RDF Query Language) is the query language proposed by W3C to query RDF for the semantic web. In syntax, SPARQL resembles SQL to a great extent (with commands like SELECT, WHERE, ORDERBY) but the difference is SQL works on a relational database while SPARQL works on the RDF files. A simple query for printing the title of a document (Figure 6-3) with keyword "knowledge management" from RDF documents is shown in Figure 6-4.

Starting from the top, PREFIX is the namespace for the URL defined in RDF. A prefix is simply a short name which can be used instead of typing the full name each time
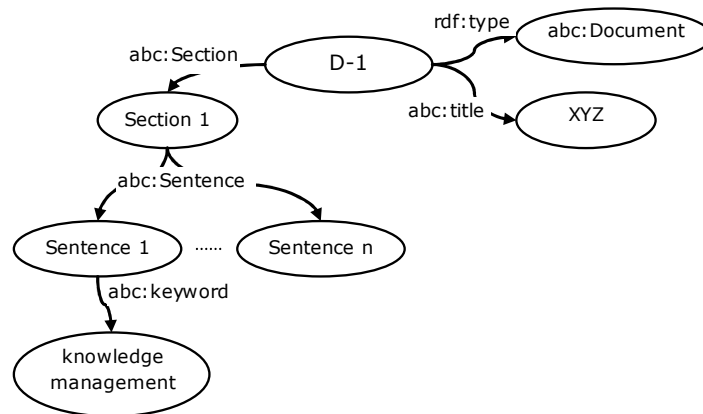
Figure 6-3: Sample RDF Document Graph

```
PREFIX abc: <http://www.yourURL.com/name/1/#>
SELECT ?Title
FROM <http://www.yourURL/doc/1.rdf>
……
FROM <http://www.yourURL/doc/n.rdf>
WHERE {
  ?x abc:keyword "knowledge management".
  ?x abc:document ?Title;
}
```

Figure 6-4: SPARQL Query for RDF Graph shown in Figure 6-3

the   URL   is   referenced.   For   instance,   abc:title   signifies   the   URI
http://www.yourURL.com/name/1/#title. The SELECT specifies the variable names
which are to be printed in the final result. Note that SPARQL variables are prefixed
with either "?" or "$". FROM is used to specify the database. Finally, the WHERE part
of the query states the pattern to be extracted from the RDF. In Figure 6-4, the first
pattern in the WHERE part finds the keyword with value "knowledge management"
and the second pattern prints the title of the document, respectively.

97

6.3.2. *Related Literature*

It is apparent that information management on the semantic web requires an ontology for semantic annotation of the information resources. The ontology is considered as an important element for semantic information management and is generally defined by the domain experts. The concepts and relations defined in an ontology are used for annotating the web documents and contextualizing them in an explicit and structured manner. The literature on the semantic annotation of web resources can be divided into two main categories, namely manual annotation and automatic annotation. Most of the manual annotation works are concerned with tools for annotating the web resources and sharing the annotations. The major concerns of these works include the representation of annotation, the ease of use, the incorporation of ontologies, the design of efficient sharing methodologies and the evaluation of annotations etc (Yang 2009). Some works in this area can be found in Stevens et al (Stevens, et al. 2003), Erdmann et al. (Erdmann, et al. 2000), Handschuh and Stabb (Handschuh, Staab 2002), Koivunen and Swick (Koivunen, Swick 2001), Martin and Eklund (Martin, Eklund 1999) and Vargas-Vera et al. (Vargas-Vera, et al. 2001). Various annotation tools for producing semantic mark-ups have also been developed such as SHOE (Heflin, Hendler 2000), Protege-2000 (Noy, et al. 2001), OntoAnnotate (Handschuh, et al. 2002), MnM (Vargas-Vera, et al. 2002), and Annotea (Kahan, Koivunen 2001).

The manual annotation approach is generally a lengthy, costly and controversial process. Automatic annotation is proposed for automatically or semi-automatically creating annotations for web resources by means of machine learning or syntactic analysis. Only a few methods have been proposed in the past to retrieve explicit knowledge from unstructured text documents expressed in natural language and then annotate it with semantic mark-ups.

Li et al. proposed a computational framework to retrieve engineering ontologies from unstructured text documents, especially engineering documents e.g. textual notebooks,

drawing etc (Li, et al. 2008). Jiang and Tan proposed CRCTOL (Concept Relation Concept Tuple based Ontology learning) for mining rich semantic knowledge in the form of ontologies from domain specific text documents (Jiang, Tan 2005). Alani et al. used the WordNet, Gate and Apple Pie parser to extract knowledge from the web, populate a knowledge base and use it to generate personalized biographies (Alani, et al. 2003).

Khan and Luo used a self organized mapping (SOM) cluster algorithm to construct an ontology (Khan, Luo 2002). Elliman and Pulido also used SOM for constructing an ontology to represent the set of web pages on a web site (Elliman, Pulido.JRG. 2002). The approach is used to construct a representation of the information required to support intelligent and interactive searching. Yang presented a semantic extraction method using a text mining approach to automatically generate some descriptive metadata and tags for web resources (Yang 2009). The proposed approach used the SOM algorithm to cluster the web pages and discover relationships between the clusters. Lee et al. proposed a methodology for automatic Chinese ontology construction from unstructured text documents (Lee, et al. 2007). The TEXT2RDF approach proposed in this chapter is to some extent similar (Lee, et al. 2007), but Lee et al's work was restricted to the Chinese documents. The detailed description of the TEXT2RDF methodology is given in the next section.

6.4. TEXT2RDF: An Approach for Semantic Document Creation

The information resources in an organization are generally described in the form of text documents published as electronic documents or hardcopy. Such documents are self-contained sets of sentences delineating certain information about the company. It is necessary to develop a standard mechanism for modelling, management and exchange of information in an explicit and structured manner. This section describes the approach proposed for extracting information from unstructured web documents and transforming it into semantic web resources. The self explanatory Figure 6-5 shows the
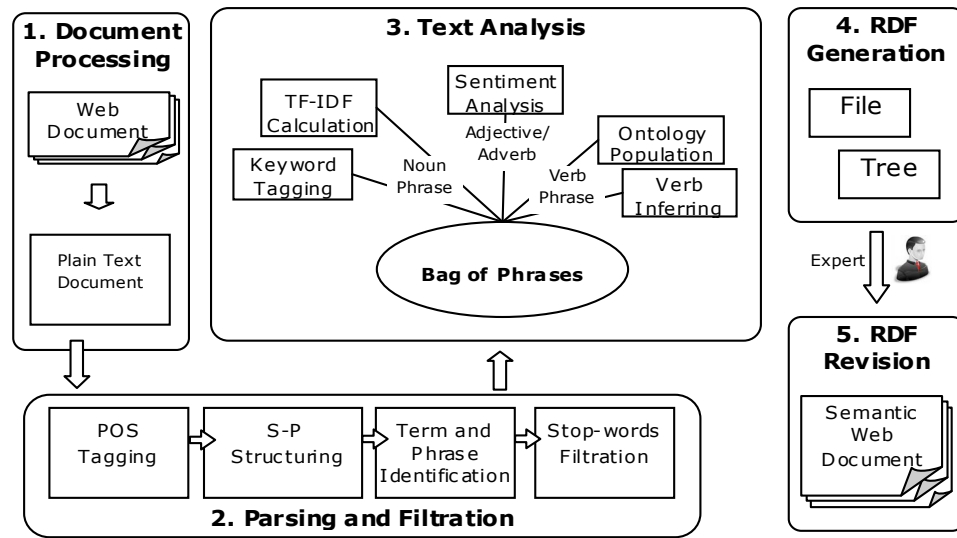
Figure 6-5: TEXT2RDF Architecture

overall architecture of TEXT2RDF proposed for extracting information from an unstructured text document. The proposed approach includes 5 processes that are described in the following subsections:

6.4.1.   *Document Processing*

Information is generally captured in the form of natural language documents such as web pages, PDF files, word documents, etc. The task of pre-processing is to convert these web documents selected manually or obtained automatically using appropriate search engine technology into a unified format. Pre-processing obtains information items from various sources and transforms them into an unstructured text document i.e. a .txt file. Note that .txt format with single headings and text in separate lines is the required type of input document used in this approach. Therefore, any documents obtained in other formats are converted to plain text documents.

6.4.2. *Parsing and Filtration*

A document is a self contained unit of discourse representing a certain point or idea. Let D be the document with a finite set of sentences $S = \{S_1, S_2, \ldots, S_n\}$ received after pre-processing. In linguistics, a sentence is considered to be a grammatical unit of one or more words $W = \{w_1, w_2, \ldots, w_m\}$ bearing minimal syntactic and semantic relation among them. Each word w connotes certain part of speech explaining how it is used in the sentence. Traditional grammar classifies words based on eight parts of speech: the verb, the noun, the pronoun, the adjective, the adverb, the preposition, the conjunction, and the interjection. The task of parsing and filtration is to label terms with corresponding parts of speech (POS) and extract phrases with words/phrases with relevant POS tags for text analysis. It is carried out in 4 steps, as shown in Figure 6-5- Step 2, i.e.:

1. *POS Tagging*: Automatic text tagging is the first and foremost step required for higher level analysis of documents in natural language processing applications. POS tagging plays an important role in many applications such as information retrieval, knowledge management, lexical ambiguity resolution, speech synthesis and resolution, machine translation etc. Automatic POS tagging has been achieved by several different approaches: rule based approach, probabilistic approaches, statistical methods, neural network approaches, and Markov chain models. The accuracy of most of these approaches is at least 95% with practically no restrictions on the input text (Church, Mercer 1993). In this research the Stanford NLP Parser is used for POS tagging (Klein, Manning 2003). It is a probabilistic parser with an unlexicalized PCFG (Probabilistic Context Free Grammars) parsing method based on the factored product model.

```
 ┌──────────────┐   IBM was founded in 1910 and is based in Armonk, New York.
 │   Sentence   │
 └──────────────┘

 ┌──────────────┐   IBM/NNP was/VBD founded/VBN in/IN 1910/CD and/CC is/VBZ
 │  POS Tagging │   based/VBN in/IN Armonk/NNP ,/, New/NNP York/NNP ./.
 └──────────────┘

 ┌──────────────┐   (ROOT
 │Tree Structure│    (S
 └──────────────┘
                 ⎧  (NP (NNP IBM))
      SUBJECT    ⎨  (VP
                 ⎩
                 ⎧    (VP (VBD was)
                 ⎪      (VP (VBN founded)
                 ⎪        (PP (IN in)
                 ⎪          (NP (CD 1910)))))
                 ⎪    (CC and)
     PREDICATE   ⎨    (VP (VBZ is)
                 ⎪      (VP (VBN based)
                 ⎪        (PP (IN in)
                 ⎪          (NP
                 ⎪            (NP (NNP Armonk))
                 ⎪            (, ,)
                 ⎩            (NP (NNP New) (NNP York)))))))
                    (. .)))
```
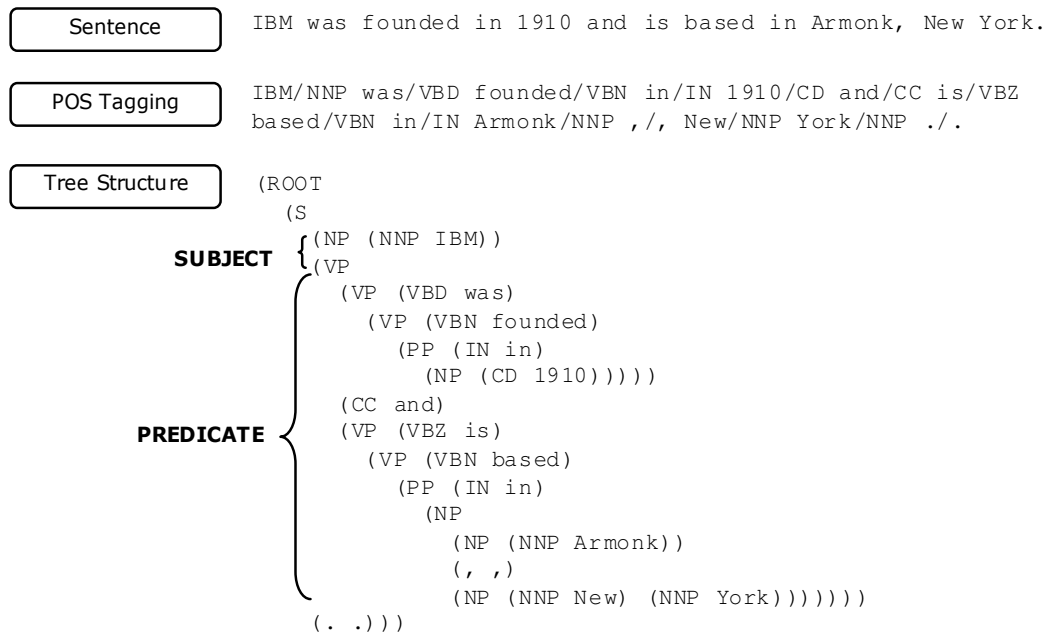
Figure 6-6: POS Tagging and S-P Structuring

The Stanford parser takes the processed sentences and constructs a tree structure for the given sentence. It uses a fixed tag set defined by Penn Treebank[1] to annotate words present in the text. For instance, a tree structure generated from the parser is shown in Figure 6-6.

2. *S-P Structuring*: According to linguistic typology, Subject-Predicate (S-P) structure is a basic pattern around which all English sentences are built. In this step, S-P structure is identified from the obtained tree structure using the rule shown in Figure 6-6. From Figure 6-6, it is evident that the first NP (noun phrase) part of tree is termed as subject and the next VP (verb phrase) part is a predicate. However, the other types of phrases (such as Adverb, Adjectives) present between NP and VP are added with VP.

---

[1] http://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

3. *Term and Phrase Identification*: In this step, key terms and phrases are identified from the S-P structure and stored in a bag of phrases (BOP). BOP is the customized form of bag of words (BOW) with a set of phrases instead of words extracted from the document. The labels used for tagging phrases include, phrase, iPhrase (important phrases, *defined in this research*), sentiment, verb, keyword, term and cardinal.

> Definition 1: Bag of Words (BOW) is defined as an unordered collection of words, disregarding grammar and even word order, such that:

$$BOW = (w_1, w_2, \dots, w_3) \hspace{3cm} \text{6-1}$$

> Definition 2: Bag of phrases (BOP) is defined as collection of phrases, such that:

$$BOP = (p_1, p_2, \dots, p_3) \hspace{3cm} \text{6-2}$$

4. *Stop words filtration*: The BOP extracted in the previous step consists of several POS that are irrelevant for further analysis (e.g. prepositions, determiners, basic english words etc). Therefore, the task of filtration is to identify relevant information from the identified phrases and pass it for text analysis. The traditional approach of filtration is to preserve certain terms (like noun, verb, adjective, cardinal number etc) and filter out other terms (like preposition, conjunction etc) depending on domain and application. This approach constructs a BOW with selected keywords and information from the document, but with a downside of losing their semantic relations and texts losing their original meaning. For instance, "developed by company" and "developed for company" are represented in the same way so the semantic relation is lost in $BOW = \{developed, company\}$. Therefore in this research, the irrelevant words are filtered by chopping Basic English words present in the front and end of selection terms/phrases. For instance, "worldwide" present in BOP is filtered with this approach and relating words such as is, "in" are retained (Figure 6-7).

IPhrase
-founded in 1910 -based in Armonk, New York

Phrase
- develops and manufactures information technology products and service

Verb
-develops -manufactures -founded -based

Sentiment
-

Term
- information technology products and services - worldwide

Keyword
- International Business Machines Corporation - IBM (2) - Armonk - New York

Cardinal
- 1910

International Business Machines Corporation (IBM) develops and manufactures information technology products and services worldwide. IBM was founded in 1910 and is based in Armonk, New York.

```
(S
  (NP
    (NP (NNP International) (NNP Business) (NNPS Machines)
(NNP Corporation))
    (PRN (-LRB- -LRB-)
      (NP (NNP IBM))
    (-RRB- -RRB-)))
  (VP (VBZ develops)
    (CC and)
    (VBZ manufactures)
    (NP (NN information) (NN technology) (NNS products)
      (CC and)
      (NNS services) (NNS worldwide)))
(. .)))
(S
 (NP (NNP IBM))
 (VP
   (VP (VBD was)
     (VP (VBN founded)
       (PP (IN in)
         (NP (CD 1910)))))
   (CC and)
   (VP (VBZ is)
     (VP (VBN based)
       (PP (IN in)
         (NP
           (NP (NNP Armonk))
           (, ,)
           (NP (NNP New) (NNP York)))))))
(. .)))
```

Figure 6-7: Text Analysis on a Sample Sentence

6.4.3. *Text Analysis*

Using the previous steps, the text document is successfully clustered in a bag of phrases. The next task is to tag these phrases with appropriate labels L, where

$$L = \{Term, Keyword, Phrase, IPhrase, Cardinal, Sentiment, Verb\} \qquad \text{6-3}$$

A set of rules are defined for identifying labels for phrases present in the document. Prior to discussing those rules, two assumptions made for these rules are as follows:

Assumption 1: A node in tree structure is a set with child nodes as elements of that set.

Assumption 2: A phrase considered for labelling does not have any similar phrase as a child node.

The rules defined for identifying labels are as follow:

- *IPhrase*: It is a list of important phrase present in a document, the phrases with keywords and cardinal terms. For example, found in 1910, based in Armonk, etc. It includes verb phrases with proper noun (NNP) and cardinal numbers in it, such that:

$$IPhrase = VP|(NNP \in VP \parallel CD \in VP)$$ 6-4

- *Phrase*: It includes all verb phrases with noun phrase (NP), excluding other types of phrases, such that:

$$Phrase = VP|(NP \in VP)$$ 6-5

- *Sentiment*: includes all verb phrases with adjective phrase (ADJP) or adverb phrase (ADVP) in it, such that:

$$Sentiment = VP|(ADJP \in VP \parallel ADVP \in VP)$$ 6-6

- *Verb*: includes verbs (VB) present in the predicate of S-P structure, such that:

$$Verb = \{VB, VBD, VBG, VBZ\} \in VP$$ 6-7

- Keyword: contains noun phrases with a proper noun in them, i.e.

$$Keyword = NP|(NNP \in NP)$$ 6-8

- Term: includes all noun phrases excluding keywords

$$Term = NP \setminus Keyword$$ 6-9

- Cardinal: It is a list of noun phrases with cardinal numbers (CD) in it, such that:

$$Cardina = NP|(CD \in NP)$$ 6-10

A simple example explaining the implementation of above rules is shown in Figure 6-7.

6.4.4.   *RDF Generation*

As mentioned earlier, RDF infrastructure is used in this research for storing documents in the semantic web. In this step, the labelled BOP is encoded into RDF using the schema defined in Section 6.2.1. An example of an RDF file generated for the text mentioned in Figure 6-7 is provided in **Appendix D**.

6.4.5.   *RDF Revision*

This is one of the important steps for semantic information management. As discussed previously, the key step in the information management is defining the possible users of that knowledge. Therefore, the RDF document is presented to an expert to define the accuracy of the semantic document generated and store it in a well accessible format and place.

6.5. Annotating Keywords with Domain Knowledge

The TEXT2RDF approach gleans information from the web documents and populates the bag of phrases. The BOP is further converted into a semantic web resource using RDF and RDF schema. In BOP, keywords represent unique entities such as name of person, organization, place etc; whereas, terms and phrases are used for representing domain specific words and concepts. These terms and keywords often differ from domain to domain and vary according to the enterprise. This hampers the interoperability and sharability of information between machines, people, enterprises etc. In order to overcome this problem, domain specific meaning needs to be added to the terms and keywords available in the document.

In this research, the semantic annotations are used for hyperlinking the terms and phrases present in documents with their meaning definitions. In general, hyperlinks are used for providing a reference or link from some point in one document to another document or database. However, in this research the hyperlinks are used for annotating the information entities with their explicit definitions provided in the local taxonomy or glossary of an enterprise. Generally, enterprises maintain their local classification systems for capturing information about terms and terminologies used in their company. These information repositories are used for adding additional bits of information to the enterprise documents and thus reducing the risk of misinterpretation while document sharing among heterogeneous information systems during the enterprise collaboration.

The proposed semantic annotation scheme is based on the following assumptions and conceptions:

Assumption 1: Information entities constitute an important part of the semantics of the documents they are mentioned in.

Assumption 2: The information entities that are not defined in the local classification systems are not of interest for the enterprise.

Conception 1: The enterprise classification systems will continuously evolve and thus new annotations will be added in the semantic document.

Conception 2: Information entities can be entities and can be coupled with formal descriptions and thus provide more semantics and connectivity to the web.

On the basis of the above assumptions, the semantic annotations are defined for terms and phrases extracted from text documents and hyperlinked with the domain specific terminologies. The annotation process is carried out by calculating the similarity between terms extracted from the documents and terminologies available in enterprise classification systems. During this evaluation, the terms are hyperlinked when they exactly correspond to a terminology defined by enterprises. Otherwise, a list of similar annotations is created and directed to the user. The detailed description of this semi-automatic process is provided in Chapter 8.

6.5.1. *Deepening of the approach*

The concept of semantic annotation of terms and keywords in web documents is used to support the idea of Linked Data. Linked data, a sub-topic of the Semantic Web, is basically about exposing, sharing and connecting data via URI (Uniform Resource Identifiers) and HTTP (HyperText Transfer Protocol) on the web (Bizer, et al. 2009). Underpinning the semantic web, it is the set of best practices for defining interconnected data, fixing standardized meanings and publishing semantic data on the web. It is used to link current web information at the data level and evolve the current

web from a global information space of linked documents to one where both documents and data are linked.

Tim Berners-Lee, inventor of the Web and proponent of the Semantic Web, outlined a set of rules for publishing linked data, paraphrased along the following lines (Bizer, et al. 2009, Berners-Lee, et al. 2001):

1.   as names for things,

2.   Use HTTP URIs so that people can look up those names,

3.   When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL),

4.   Include links to other URIs, so that they can discover more things.

These 'Linked Data principles' provide a basis for publishing and connecting data using the infrastructure of the Web while adhering to its architecture and standards. The Linking Open Data[2] (LOD) project, an open and collaborative effort by W3C SWEO Community, aimed to develop linked data by publishing open datasets as RDF and creating RDF links between data items from different data sources. The project started out in early 2007 with a relatively modest number of datasets and participants and has grown since both in terms of depth, impact and contributors. Currently LOD contains more than 40 datasets, with total volume above 4.7 billion statements, interlinked with 142 million statements as illustrated on Figure 6-8.

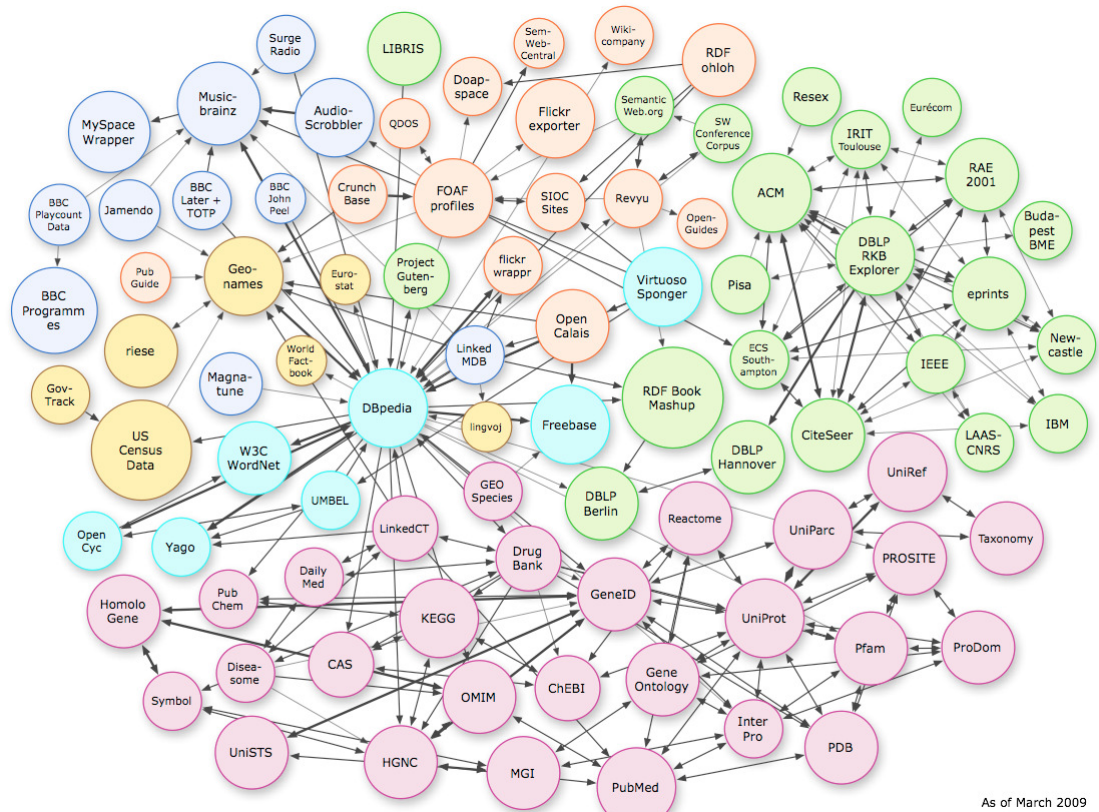
[2] http://linkeddata.org/

Figure 6-8: Linking Open Data (LOD)

The proposed concept of information extraction and semantic annotation is introduced to support the idea of Linked Data. The idea is, if enterprises can publish their glossary, directory or other relevant information with Linked Data, then enterprise documents can be linked with this explicit information and converted into a semantic document publishable and accessible over the global information space, without any risk of misinterpretation. This idea can facilitate the sharing of information resources among enterprises from different background, culture or location, or simply heterogeneous systems within one enterprise to interoperate at the data level.

6.6. Summary

The operation stage of a virtual enterprise is considered as the actual project period whereby the conceptualization to the delivery of the products or services to the customers is achieved within a promised time window. During this stage, each enterprise is involved as an autonomous entity, but exposes and shares their information resources to satisfy the requirements of partners and achieve the goals of the project. However, in a virtual networked environment information resources are inherently distributed and expressed with domain specific keywords and phrases. These domain specific terminologies often vary according to the enterprises and thus hamper the interoperability and sharability of information between enterprises etc. In order to support information sharing and communication between heterogeneous enterprises, it is necessary to describe information resources in a way that is understandable and usable by the networked organizations.

In this chapter, a TEXT2RDF application is proposed to extract terms and phrases from the enterprise document and annotate them with domain specific terminologies of the enterprises. This text mining application automatically extracts information entities from the text documents and attaches an extra piece of information to them. This additional information is the meaning and definitions of concepts used in the documents. In this research, the concepts definitions available in the local classification systems of the enterprises, such as directories, catalogue and yellow pages are considered for semantic annotation of documents. In future, if the enterprises publish their terminologies on the web or link with LOD (Linking Open Data), then the proposed semantic annotation concept can be exploited to create a semantic web document annotated with globally accepted and accessible definitions.

C h a p t e r   7

TERMINATION STAGE: MANAGING ENTERPRISE KNOWLEDGE

The knowledge economy is the economy of the future
Tony Blair, 1999

ABSTRACT

Termination stage is the last stage in the life cycle of virtual enterprise, when the collaborative project is finished and the network is disbanded. During this stage, it is necessary to gather knowledge and expertise from project members and utilize it in future projects. One of the central themes of knowledge management is the design, building and maintenance of an effective corporate memory. Due to its continual importance and popularity among enterprises, a framework for corporate memory management on the semantic web is proposed in this chapter.

CHAPTER OUTLINE

7.1. Introduction

A virtual enterprise is regarded as a rich open information space with large amounts of shared information and knowledge, ranging from practical to conceptual and from narrow to broad, accumulated over the entire life cycle of the virtual enterprise. The information is generally stored in a shared resource repository, consisting of massive information resources (e.g., electronic documents, multimedia data and so on) and processed and managed by multiple information systems providing relevant services and functionalities (Jonkers, et al. 2004). However during the termination stage of a virtual enterprise, the shared information is no longer needed in its shared form. At this final stage, the project related information is archived and distributed among partners; to learn lessons and utilize it in future projects. It is vital that the new and shared knowledge from the collaboration is dealt with in accordance with any collaboration agreement and to maximise the commercial benefits of the partners.

The knowledge and expertise gained from collaboration is arguably the most valuable asset for any organization. This information is like a raw material, which when processed quickly and cost effectively, can be used for value addition. But, if not filtered, refined, delivered and processed properly, the information becomes practically useless (Wick 2001). This observation stresses the requirement for employing knowledge intensive methodologies that facilitate meaningful acquisition, sharing and reuse of information as a main source of competitive advantage for corporations. In organizations, knowledge management is now recognized as an essential strategy for ensuring that the right information is delivered to the right person at the right time.

Knowledge management is the process that takes an organization's data and information and turns it into knowledge for tangible benefits for the company. It captures intellectual assets by converting data, organizational insight, experience and expertise into reusable and useful knowledge that is distributed and shared with the people who need it. Knowledge management includes acquiring, cataloguing, storing,

and distributing of data, information and knowledge. In the context of knowledge management (KM), an explicit and disembodied representation of knowledge is known as *corporate memory* (Van Heijst, et al. 1996). Corporate memory (CM) is the total body of data, information and knowledge required to deliver the strategic aims and objectives of an organization. It explicitly stores collective knowledge possessed by human capital and technology, organizational structures and organizational culture; and makes it available to the entire company. The major benefits of this explicit storage are: reduced risk of loss of analytical knowledge, better knowledge flow, and reduction in the time and demands placed by human capital (Huang, et al. 2005).

Although, knowledge storage and maintenance within a corporate environment have been considered for decades, it has always been an expensive and risky proposition (Lai 2007). It is only with the advent of web technology that information sharing has emerged as a technology area. The Web has enabled firms to make their information available electronically so that users can access and share their resources and expertise electronically. Despite growing interest and efforts, web technology is still primitive in its functionality. Today's web arranges the information syntactically, which means most information has to be interpreted by humans before use, rather than being processed automatically by machines. To overcome this problem, researchers are focusing on semantic web concepts and tools that enable computers to automatically process and understand the information (Casey, Pahl 2003). The primary benefit of this new vision is to represent web resources in formalisms that both machines and humans can understand.

Keeping the present technological scenario in mind, this research aims to:

Develop a framework for representing corporate memory on the semantic web and identify relations between the semantic documents.

The idea is to provide explicit semantic descriptions for corporate memory and encode it in an unambiguous and machine readable form. The proposed framework

mainly consists of two processing steps: firstly the unstructured text documents present in corporate memory are converted into a semantic web resource using the proposed TEXT2RDF application. The TEXT2RDF application proposed in Chapter 6 is used to glean information from documents and converts it into a semantic web resource using RDF and RDF Schema. Subsequently, relations need to be identified between terms and phrases present within documents and among other semantic documents available in corporate memory, to facilitate effective knowledge reuse. The proposed framework with information extraction and semantic annotation is used to facilitate following applications in a virtual collaboration (Missikoff, et al. 2003):

- *Semantic Interoperability*: Annotating information resources with local semantics to support business cooperation among enterprise software applications.
- *Web Services publishing and discovery*: Semantic matchmaking of information producer and information consumer.
- *Document Management*: Assist in navigating and analysing the set of documents by Identifying relations between them.
- *Knowledge Management*: Organization and retrieval of enterprise knowledge;

The implementation of the first two applications was shown in the operation stage of the virtual enterprise (in Chapter 6), whereas the last two applications are realized during the termination stage of the virtual enterprise, discussed in this chapter. The rest of this chapter is organized as follows: Section 7.2 provides brief details about corporate memory and related literature on corporate memory management. Section 7.3 gives details on how documents in corporate memory are related and queried. Finally, the chapter is summarized in Section 7.4.

7.2. Corporate Memory: Strategy for Enterprise Knowledge Management

Knowledge and expertise are the most valuable assets for any organization. Effective organizations are those that recognize the process of mapping out where these assets reside and identifying the conditions that foster their generation and re-use. In the

current economy, investment in knowledge assets is considered far more profitable than in any materialistic assets. With knowledge management, organizations seek to create an environment where knowledge is accessible to everyone regardless of their format, requirements or physical locations. Knowledge management is a discipline that provides strategy, process, and technology to share and leverage information and expertise for improved performance, competitive advantage and continuous improvement of the organization. One of the central themes of knowledge management is the design, building and maintenance of an effective corporate memory (Van Heijst, et al. 1996). A corporate memory is a resource for preserving valuable heritage, learning new things, solving intricate problems, creating core competencies and initiating new situations for both individuals and organizations now and in the future.

7.2.1. *Corporate Memory*

The concept of corporate memory has been defined by several researchers in different ways but one that is widely referenced and appropriate for this research is (Rabarijaona, et al. 2000):

> Explicit, disembodied, persistent representation of knowledge and information in an organization, in order to facilitate its access and reuse by members of the organization, for their tasks.

Corporate memory is a major asset in any organization. It mainly consists of i) knowledge possessed by people and technology, ii) organizational structure and iii) organizational culture. In order to create and maintain a corporate memory, it is essential to plan and establish a workable record keeping system. According to (Dieng, et al. 1999), the formation of corporate memory relies on the following steps:

- Detection of needs: The first and foremost step is to identify the needs of corporate memory, such as who are the users, which tasks they have to perform, in which situations, which knowledge types they need to memorize and retrieve, which tools they use, etc.

- Construction: A corporate memory includes all kinds of written documents, such as: reports, notes, newsletters, contracts, licenses etc. The information must follow a useful format to make things easier for others.

- Diffusion and use: The corporate memory is useful, if it is well kept and accessible. Therefore, adequate elements of memory must be distributed to the relevant members of the organization.

- Evaluation: Corporate memory is expensive and space consuming (computer memory or space) in terms of storage. Therefore its evaluation for appropriate selection is important, from several viewpoints such as economico-financial, socio-organizational and technical.

- Maintenance and evolution: In order to get benefits from this knowledge store, it must be maintained and evolved with high priority.

Corporate memory has been a hot topic in management and computer science, at least since the early 1990's. The next subsection gives a brief review of the published work in the field of corporate memory.

7.2.2. *Related Work*

Van Heijst et al. presented some initial thoughts on how corporate memory should be organized in order to achieve maximum contribution in the competitiveness of an organization (Van Heijst, et al. 1996). Rabarijaona et al. showed the advantages of XML meta-language and ontology for corporate knowledge management (Rabarijaona, et al. 2000). In (Huang, et al. 2005), XML is implemented to design the structure of knowledge and construct a standard corporate memory according to the characteristics of different data-analysis techniques. A XML transformation process is proposed in their paper for converting a general document into an XML based document; and this is validated on a manufacturing case study. Verma and Tiwari showed an XML based representation of corporate memory for supplier selection in a global supply chain problem (Verma, Tiwari 2009).

The former works have utilized different web based tools and languages for the representation of corporate memory. The purpose of introducing web technology in corporate memory management is to enhance the effectiveness of communication and management necessary for its capture and reuse. However, even if information is easily accessible it may still not be straight forward to use or fully understand. Current web tools provide no information about the semantics of the described language. It is dependent on the individual programmer or interpreter to understand the semantics of the language from the available description. In order to overcome this drawback, it is necessary to make knowledge explicit and machine interpretable.

In the age of the knowledge economy, corporate memory needs sets of interconnected data and semantic models to communicate and share information. El-Diraby and Zhang 2006 developed a semantic framework for representing and utilizing corporate memory in the building construction domain (El-Diraby, Zhang 2006). This framework is based on a taxonomy with 6000 concepts arranged with explicit definitions and interrelationships. Rios-Alvarado 2009 introduced ODARyT, a semantic web approach to represent and retrieve information in a corporate memory (Rios-alvarado, Medina-ramírez 2009). The tool developed in this research is used for editing and browsing an ontology, using the ontology to annotate heterogeneous resources, and compose a corporate semantic web.

7.3. Framework for Corporate Memory Management

Enterprises with significant corporate knowledge generally create an environment to facilitate its better reuse and deployment in decision-making processes. The corporate memory management is an infrastructure which is guarded and grown to provide technical support for enterprise knowledge management. According to Kuhn and Abecker, corporate memory management is a comprehensive computer system that captures a company's accumulated know-how and other assets to improve the efficiency and effectiveness of activities (Kuhn, Abecker 1997). Linked with knowledge

management tasks, corporate memory is used to capture information from various sources of an organisation and make it available for different tasks. D. Grey stated the purpose of corporate memory management as (Grey 2003):

> One of the central themes of KM is the design, building and maintenance of an effective 'corporate memory', a repository, a dare I say it, knowledge-base. Here the intellectual jewels of the organization will reside, easily accessible, expertly indexed, intuitively browsable. Here experts and novices will come for self-help knowledge, they will find the correct solution quickly, be able to apply the solutions with confidence, and learn from the 'collective experience of the organization'.

Considering the above requirements, this research proposes a framework for managing corporate knowledge in an enterprise. The motivation for the proposed framework is to improve the semantic understanding of corporate knowledge resources. The semantic web is a key technology for achieving better management of knowledge that makes information accessible and understandable not only by humans but also by computers, thus enabling an extended cooperation among people and machines (Berners-Lee, et al. 2001). In the context of the Semantic Web, there is a meeting point between Web technology and corporate knowledge: both gather heterogeneous and distributed information and share the same concern about the relevance of information retrieval. However, corporate memory has a context, an infrastructure and a scope limited to the organization where they are applied (Rios-Alvarado Ana, et al. 2009).

In this chapter, a framework is developed to represent and retrieve knowledge from heterogeneous information resources, identify relations among them and utilize it in various enterprise activities such as learning, information interchange, improve information retrieval and information distribution. The proposed framework focuses on document based information and consists of four steps as shown in Figure 7-1 and explained as follows:
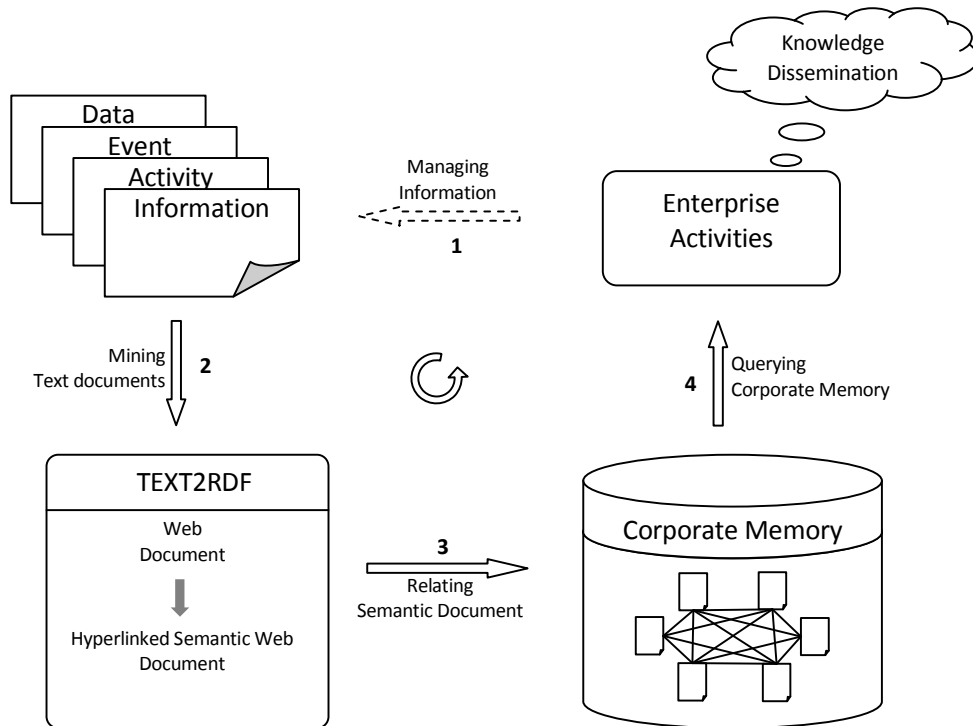
Figure 7-1: Framework for Corporate Memory Management

7.3.1. *Managing Information*

An enterprise of any kind, e.g. an actual company or a virtual enterprise with distributed companies working together online, deals with the continuous growth of heterogeneous information resources. This information consists of (Huang, et al. 2005):

- general information, described as fundamental information in the corporate memory, for example, subject-oriented information, emails, notes, etc;

- technical information, which includes information (data) analytical technology and information from a data-storage system for example, manuals, guides, application notes, datasheets;

- captured information which is derived from information (data) analytical technology without receiving any content from experts, for example, analytical events, analytical topics, classified information;

- refined knowledge which integrates captured and technical information with subject-oriented expertise, such as feedbacks, reports etc.

The above mentioned corporate information, in general is extracted from data storage systems or domain experts, and may have varying formats and locations e.g. non computational, database-based, document-based, knowledge-based, case-based, Web-based resources. However, the proposed framework focuses on document based corporate memory, distributed through the Web.

7.3.2.   *Mining Text Document*

Several kinds of documents can be identified in a document based corporate memory e.g. project documents, technical reports, reference manual, emails etc. The text mining approach, TEXT2RDF, proposed in the previous chapter is used to extract basic information from the documents available in different formats and convert it into a semantic web resource using RDF and RDF Schema. This computer understandable document is then stored in a semantic database (e.g. RDF triple store) and information is extracted from it using SPARQL query language. The terms and phrases identified in the semantic documents are further linked with domain specific knowledge and a wiki-like document is created using the enterprise glossary and/or terms available on the web.

The hyperlinked semantic web documents offer the following advantages for corporate memory management:

- *Information exchange standard*: In an enterprise, information streams in an unformatted and raw structure and thus cannot usefully be stored in its original form. It is essential to convert it into a stable and explicit form and use it for facilitating communication and exchange of information. In this research, RDF/RDF-schema is used as a standard format for capturing and sharing semantics of information in a standardized interoperable manner.

- *Managing heterogeneous documents*: A corporate memory might include documents in varying formats e.g. web pages, PDF files, word docs etc. The TEXT2RDF approach converts heterogeneous documents into a standard format and also provides a standard metadata for it.

- *Multiple view of the same information*: The TEXT2RDF approach uses RDF as a standard format for converting the text documents as a semantic web resource. As discussed in Section 4.3.4. RDF documents can be stored and presented in different formats, i.e. graph based, XML syntax or N-triple. In this research, RDF/XML format is used for presenting the semantic documents. RDF/XML syntax supports information presentation in a uniquely structured manner, which separates the documents logical structure from its presentation and thus a style-sheet can present a document as HTML, PDF or any other format.

- *Annotation with metadata*: The TEXT2RDF approach annotates a document with metadata that describes the document abstractly and synthetically, according to predefined standard ontologies. A user can use this metadata to establish relations with other ontologies and knowledge models. Further, search engines and knowledge servers can also use these semantic tags to facilitate their processes.

- *Hyperlinked with domain specific knowledge*: The hyperlink process adds external links to the documents from the outside without modifying the source. This hyperlinked document can be used to identify relationships between terms and phrases present within a document and among other semantic documents available in corporate memory.

- *Information search*: RDF/XML facilitates the information search because documents are structured and hence can be considered as a database. A standard mechanism, SPARQL in this research, can be used to search the information in a structured way.

7.3.3. *Relating Corporate Memory*

The next task in the proposed framework is to relate semantic documents by identifying relationships between terms and phrases extracted from the documents. In this research, terms and phrases are handled separately and differently for identifying relations with other phrases in a corporate memory. The approach applied for relating documents in a semantic corporate memory is explained as follows:

7.3.3.1. *Keyword and Term*

The hyperlink process in the TEXT2RDF approach is used to link terms and keywords to a single point in a business directory or dictionary. The words labelled as keywords can be selected from the documents available in corporate memory and hyperlinked with a single URI. However, the words tagged as *term* impose severe constraints in finding relations. Organizations often use 1) multiple words that have similar meaning (synonym), and 2) words with more than one meaning (polysemy). These two constraints cause mismatches in the vocabulary used in corporate memory and information retrieval systems. In order to overcome this drawback the following approaches are being developed by researchers to integrate semantic processing into their information storage and retrieval system, e.g. auxiliary structure, local co-occurrence statics and latent semantic analysis (Bradford 2008).

Latent Semantic Analysis (LSA) is one of the most promising tools used for finding patterns of relationships among a collection of documents. It is an indexing and retrieval method capable of identifying patterns of relationships between concepts, not just matching specific keywords. A key feature of LSA is its ability to extract the conceptual content from text by establishing associations between terms that occur in similar contexts. It has been used in a variety of information retrieval and text processing application such as, text summarization, information discovery, online customer support, information visualization, etc (Bradford 2008). LSA relies on Singular Value Decomposition (SVD) of a matrix (term-document) to determine relationships between

terms and concepts used in the documents. In this research, LSA is carried out in 4 steps on a set of corporate memory documents and words are labelled as terms in those documents:

1. *Term-Document Matrix*: Let $D = \{d_1, d_2, \ldots, d_n\}$ be the set of documents present in corporate memory and $T = \{t_1, t_2, \ldots, t_m\}$ be the set of words tagged as term extracted from the documents. Initially, a term-document matrix $M_{m \times n}$ is constructed such that $M = [m_{ij}]$ where $i = 1, 2, \ldots, m;$ $j = 1, 2, \ldots, n;$ $m_{ij} = frequency(t_i, d_j)$.

2. *Singular Value Decomposition*: The matrix $M_{m \times n}$ is then subjected to SVD, which decomposes the M into the product of 3 other matrices, such that:

$$M_{m \times n} = USV^*$$ 7-1

Where, $U_{m \times m}$ = row entities as vectors of derived orthogonal factor values

$V_{n \times n}$ = column entities as vectors of derived orthogonal factor values

$S_{m \times n}$ = diagonal matrix containing scaling values such that when three matrices are multiplied, $M_{m \times n}$ is reconstructed.

3. *Dimensional Reduction*: In this step, all but the *k* (here *k*=3) largest values are set to 0. This dimensional reduction leads to a matrix M_k such that

$$M_k = U_k S_k V_k^*$$ 7-2

where, U_k, S_k, V_k have dimensions $m \times k$, $k \times k$ and $n \times k$ dimensions respectively.

4. *Graphing Document-Term*: Finally, the terms and documents are plotted in a $X - Y$ graph with $U_{m \times k}$ and $V_{n \times k}$ respectively. Here the matrix values with $k = 2$ and $k = 3$ are taken as xy coordinates in the graph. It places both terms and documents in the same graph and thus helps in identifying clusters of terms and documents.

Here is a simple and illustrative example explaining the above LSA approach:

Suppose a set of four text documents, two from financial sector and other two from IT (information technology) sector. In order to demonstrate the analysis step by step, a

Table 7-1: Document Term Matrix

	fund	credit	training	company	computer	networking	e-commerce
F1	2	1	1	1	1	0	0
F2	0	2	1	1	0	0	0
I1	0	0	1	1	1	1	1
I2	0	0	0	1	2	1	0

limited set of keywords is taken from each document. The first step in LSA is to create a word by document matrix. The matrix $M_{7\times4}$ with 7 keywords and 4 documents is shown in Table 7-1. In this matrix, each keyword is a row and document is a column. Each cell contains the number of times that word occurred in that title. For example, the word "computer" appears once in F1 and I1 and twice in I2. In general, these matrices tend to be very large and sparse with most cells containing 0.

Once the document-term matrix is built, the SVD is applied for analyzing the matrix. The SVD is used to find a reduced dimensional representation of $M_{7\times4}$ that emphasizes the strongest relationships and throws away the noise. The three decomposed matrices U, S and V^* is as follows:

$$U = \begin{matrix} -0.631 & 0.336 & -0.689 & -0.121 \\ -0.43 & 0.585 & 0.631 & 0.274 \\ -0.456 & -0.383 & 0.357 & -0.72 \\ -0.458 & -0.631 & 0.002 & 0.626 \end{matrix}$$

$$S = \begin{matrix} 3.881 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.43 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.664 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.123 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.987 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.674 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.234 \end{matrix}$$
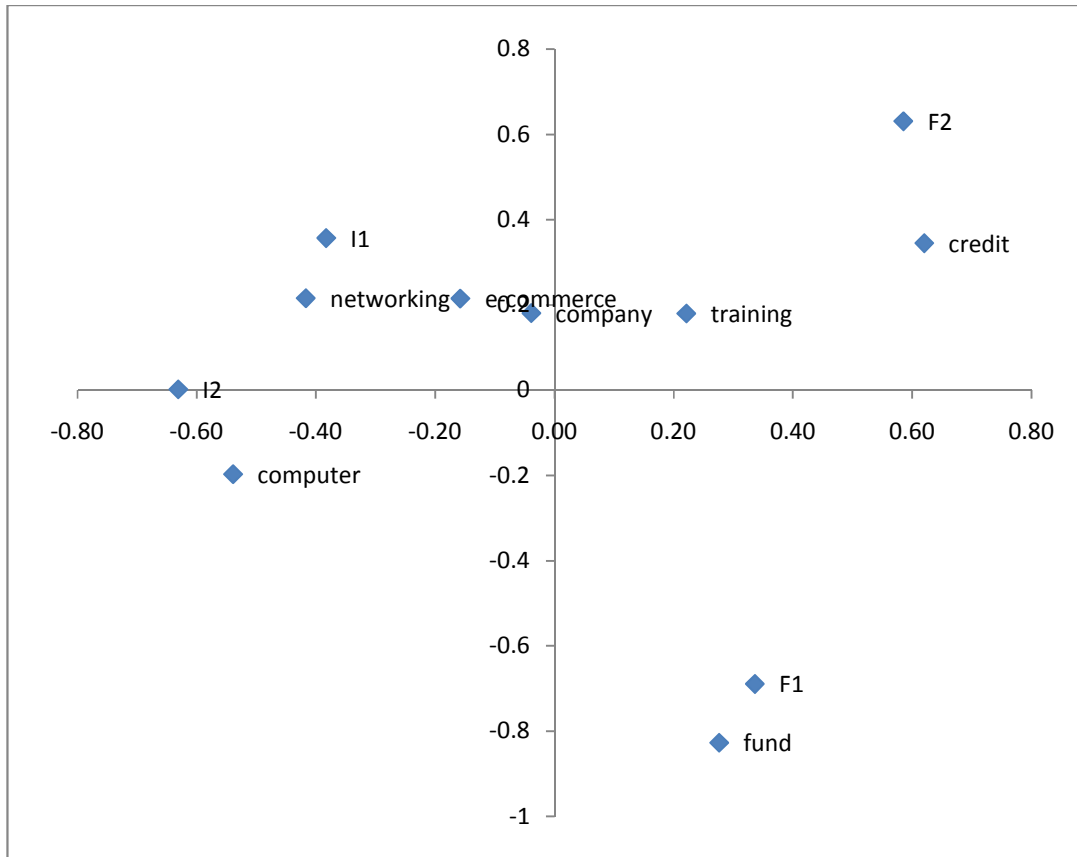
Figure 7-2: Document Term Graph for matrix shown in Table 7-1

$$V^* = \begin{matrix}
-0.325 & 0.276 & -0.827 & -0.215 & -0.265 & 0.124 & -0.037 \\
-0.384 & 0.62 & 0.345 & 0.38 & -0.276 & 0.085 & 0.35 \\
-0.391 & 0.221 & 0.18 & -0.505 & 0.671 & 0.246 & 0.02 \\
-0.509 & -0.039 & 0.181 & 0.053 & -0.118 & -0.416 & -0.719 \\
-0.516 & -0.539 & -0.197 & 0.367 & 0.253 & -0.163 & 0.424 \\
-0.235 & -0.417 & 0.216 & -0.083 & -0.388 & 0.742 & -0.128 \\
-0.117 & -0.158 & 0.215 & -0.641 & -0.418 & -0.409 & 0.403
\end{matrix}$$

Subsequently, the three matrices U, S and V^* are dimensionally reduced and first k columns of each matrix are considered for further analysis. In this example, the matrices are reduced to $k = 3$, the first column is thrown and second and third column are used to build the graph. The reason first column is not considered because it correlates with the number of times that word has been used in all documents. It can be considered for analysis by centring the matrix, but it will dramatically increase the memory and

125

Table 7-2: Distance measure between documents and terms

	F1	F2	I1	I2
fund	0.022644	2.221245	1.836137	1.836137
credit	1.149812	0.083021	1.006153	1.006153
training	0.768386	0.335897	0.396145	0.396145
company	0.897525	0.591876	0.149312	0.149312
computer	1.007689	1.94896	0.331252	0.331252
networking	1.386034	1.176229	0.021037	0.021037
e-commerce	0.718757	0.725105	0.070789	0.070789

computation requirements of the analysis. Therefore, the second and third column of U, S and V^* is considered for building a document-term graph, shown in Figure 7-2.

The main advantage of this graph is that both terms and documents are plotted in same graph. Thus, it can be used to identify the cluster of terms and documents closely related to each other. Further, it shows that the terms occurring in any document lie closer in the graph in comparison to the other terms in the set of terms. As evident from Figure 7-2, the terms lying closer to the document nodes are occurring in matrix $M_{7\times4}$. The distance between documents and terms provided in Table 7-2 also clarify the closeness of terms with documents.

7.3.3.2. Phrase and IPhrase

The similarity between two phrases or iphrase is calculated using Dice's coefficient, defined as:

$$Sim(phrase_1, phrase_2) = \frac{2 \times CommonTerm(phrase_1, phrase_2)}{NumberOfTerm(phrase_1) + NumberOfTerm(phrase_2)} \qquad 7\text{-}3$$

where, $Sim(phrase_1, phrase_2)$ is the Dice's coefficient (Dice 1945), $NumberofTerm(phrase)$ indicates the total terms in a phrase and $CommonTerm(phrase_1, phrase_2)$ refers to the number of common terms used in $phrase_1$ and $phrase_2$.

7.3.3.3.   Verb

In contrast to the above two approaches used for relating terms and phrases among documents, the verb tags are used for finding relations and clustering sentences in the same document. In linguistics, a verb is considered to be the most important part of sentence linking two nodes (Subject and Object) in a sentence. The terms labelled with verb tags are classified using the predefined classes defined in Verbnet[3]. Verbnet is a hierarchical verb lexicon with syntactic and semantic information for English verbs using Levin verb classes to systemically construct lexical entries (Dang, et al. 1998). Currently, Verbnet has a list of over 5200 verbs classified in 270 top classes. By using the Verbnet classes, the sentences present in a document can be easily clustered and classified in different groups. For instances, three groups of verbs used in a sample document can be:

* *Create-* create, manufacture, product, develop, construct, etc
* *Require-* require, need, demand, involve, etc
* *Use-* use, employ, utilize, apply, etc.

In real time, enterprises can use these verb classes and define their own top level classes for their organizational corporate memory.

7.3.3.4.   Sentiment and Cardinal

The words labelled as sentiment and cardinal are phrases providing document specific information. Therefore, these terms are only highlighted in the document and stored separately in semantic web resource.

7.3.4.   *Querying Corporate Memory*

The semantic corporate knowledge with hyperlinked text and semantic relations with other documents in a corporate memory can be used to expedite various enterprise activities. It can promote knowledge growth, knowledge communication and knowledge

---

[3] http://verbs.colorado.edu/~mpalmer/projects/verbnet.html

preservation in organizations. In order to use this semantic knowledge, knowledge servers can retrieve information from these structured documents using SPARQL query language, discussed earlier in Section 6.3.1.

7.4. Summary

Termination stage is the last stage of virtual enterprise life cycle, when the project has finished its task by delivering the required product/services and collaboration is no longer needed. During this stage, the shared knowledge is no longer required in its shared form. Each partner has gained knowledge through its experience and learning as part of the virtual enterprises. At this stage, the need is to develop mechanisms for capturing lessons learnt from knowledge.

To achieve and exploit the benefits available from knowledge resources, knowledge management techniques are implemented to share and leverage information and expertise for improved performance, competitive advantage and continuous improvement of the organization. One of the central themes of knowledge management is the design, building and maintenance of an effective corporate memory. Corporate memory is the total body of data, information and knowledge required to deliver the strategic aims and objectives of an organization. Due to its continual importance and popularity among enterprises, a framework for corporate memory management on the semantic web is proposed in this chapter. This framework mainly consists of two processing steps: firstly the unstructured text documents present in corporate memory are converted into a semantic web resource using the proposed TEXT2RDF application. Subsequently, relations are identified among semantic documents available in corporate memory.

The implementation of proposed framework on an experimental case study is provided in Chapter 8.

C h a p t e r   8

CASE STUDY

There comes a time when every scientist, even God, has to write off an experiment.

P. D. James

ABSTRACT

This chapter presents a case study to demonstrate the prototype implementation of the contributions made in this research. A real life case study of a virtual enterprise is taken to illustrate the applicability and effectiveness of the semantic tools and techniques proposed for each stage of the virtual enterprise.

CHAPTER OUTLINE

8.1. Experimentation Setup

The primary motivation behind this chapter is to provide a context for presenting a real life implementation of the semantic frameworks which has been proposed and introduced in Chapter 4-7. This prototype implementation is based on a real life case study of a virtual enterprise and is included to demonstrate and analyse the feasibility and applicability of concepts proposed for each stage of the virtual enterprise. This exemplar case study presents the implementation and evaluation of the semantic tools and techniques, which are previously proposed in this thesis.

In this research, a Virtual Enterprise is considered to be a temporary alliance of independent enterprises coming together to share skills, core competencies and resources in order to better respond to business opportunities. Before a virtual enterprise can be created, there needs to be a collaboration pool of enterprises that are willing to work together, if a suitable opportunity arises. A collaboration pool provides a basis for social relationship and trust among enterprises and thereby facilitates and fosters future collaborative business ventures. The enterprises involved in the collaboration pools publish information about their competences and utilize it to envisage and plan the execution of new projects, initiate new business opportunities, and extend their business potential. Once an appropriate tender is identified and partners are selected for a project, enterprises start working together with shared goals and specific tasks. Finally at some point when the goals of the virtual enterprise are achieved, the enterprises disband and should archive their relevant knowledge and learning for future projects.

In this research, semantic tools and techniques are proposed to facilitate information sharing and knowledge management throughout the entire life cycle of a virtual enterprise. A case study is presented in this chapter to illustrate the applicability and effectiveness of the technologies proposed for each stage of the virtual enterprise life cycle. The experimental setup for the case study is as follows:

Consider a collaboration pool with 25 enterprises, each having varied competences, skills and processes, and a willingness and interest to collaborate if a suitable business opportunity is identified. The step wise processes followed to demonstrate each stage of virtual enterprise life cycle are as follows:

- *Identification Stage*: At first, the basic information about the 25 enterprises are collected from the web and captured and formatted using the ECOS ontology. In this research, information has been collected and analyzed from the web pages of real companies, but then bearing in mind the copyright issue, the name of enterprise, their customers and basic details were changed manually. The enterprise information has been converted into a semantic web resource using the ECOS-form and an ECOS-card is created for each enterprise.

- *Creation Stage*: At this stage, the ECOS-cards created for each of the enterprises are collected together and a collaboration pool of enterprises has been formed. The ECOS-cards are then analyzed and compared to identify direct and indirect relationships among the enterprises. Further, these ECOS-cards are evaluated against business opportunities and based on the business requirements of the opportunity appropriate business partners are identified from the members of the collaboration pool in order to create a virtual enterprise. The example business opportunities are captured from an existing tender-alert service, i.e. *Sell2Wales*.

- *Operation Stage*: A virtual enterprise with partners from three different sectors is considered as an example for this stage of virtual enterprises. Based on the assumption that enterprises perform their own tasks and share information over the computer networks with other partners, the proposed TEXT2RDF application is implemented on text documents shared among them. A semantic glossary of terminologies is developed for different sectors of enterprises and is used for annotating the terms extracted from text documents using TEXT2RDF application.

- *Termination Stage*: A corporate memory framework is proposed for the termination stage of virtual enterprise, which stores text documents in explicit and machine understandable manner and identifies relationships between them. Two types of corporate memory documents, namely professional memory and project memory, are considered to show the feasibility and effectiveness of the proposed framework. Initially, TEXT2RDF application is implemented for extracting keywords and phrases from the document and then relationships are identified between the documents using the proposed corporate memory framework.

The detailed description of proposed case study and implementation of research ideas are provided in following sections and can also be seen on web http://kmm.lboro.ac.uk/projects/phd#NiteshKhilwani.

8.2. Prerequisite work

This section presents the approach used for the implementation of the proposed frameworks and development of the prototype web applications. Generally, the web applications are designed and developed using the layered architecture, in which the

user interface, application processing and data storage are developed and maintained as the independent modules (Eckerson 1995). The advantage of the layered architecture is that each module is handled as a logically separate process, and therefore modification in one layer does not imply a modification in another layer. Further, it guarantees maximum clarity, easy debugging of errors, and code reuse.

In this research, the prototype application is developed using the *three layered architecture*, with Presentation, Application and Data management layers. The *presentation layer* involves presentation of the information, interaction with the users and communication with the application layer. The *application layer* is mainly used for executing the technologies on the information and then communicating the processed information to the other layers for its presentation or storage. The *data management layer* is responsible for managing the information and responding the requests from applications layer.

The entire prototype web application developed in this research is based on the abovementioned three layered architecture and implemented in a JAVA based environment. The Java based technologies used in each layer of prototype architecture are as follows:

- *Presentation* Layer: The entire presentation of the prototype web application available on the web (http://kmm.lboro.ac.uk/projects/phd#NiteshKhilwani) and framed as screenshots in later sections, is developed using GWT (Google Web Toolkit). GWT[1] is an open source set of tools used for creating and maintaining complex front-end web applications in Java.

- *Application layer*: The semantic web frameworks proposed for each stage of the virtual enterprise life cycle is developed and executed in this layer. All the

---

[1] http://code.google.com/webtoolkit/

applications are coded in Java language using Java NetBeans 6.7.1 IDE[2] and executed on a Toshiba Personal Computer with Pentium Central Processing (CPU) at 2.16 GHz. These applications utilize various open source software packages, such as Jena API[3], Stanford NLP Parser[4], Treebolic[5], etc. The details of these software packages are provided in **Appendix E**.

- *Data management layer*: In this research, the focus is mainly on the semantic data generated in the application layer. This semantic data is mainly in the form of RDF documents available in RDF/XML format. These RDF documents are mainly stored in *triple store*, a purpose-built database developed for managing RDF documents. In this research, Jena TDB[6] a triple store from Jena API is used for the storage and retrieval of the semantic data (Details in **Appendix E**).

The next section presents the implementation of above architecture on a real life case study.

---

[2] http://netbeans.org/community/releases/67/

[3] http://jena.sourceforge.net/

[4] http://nlp.stanford.edu/software/lex-parser.shtml

[5] http://treebolic.sourceforge.net/en/index.html

[6] http://openjena.org/TDB/

8.3. Identification Stage

During the identification stage of a virtual enterprise, companies come to the open market to share and effectively utilize their skills, core competences and resources to gain business. To explore opportunities in the global market, enterprises generally publish their skills, capability, expertise and competence that they are able to offer. ECOS is proposed in this research as a standard ontology for representing enterprise competence and publishing it as a semantic profile on the web. This ontology model captures the published competences of an enterprise using the ECOS-form and presents it as a semantic web resource, called an ECOS-card. The main feature of ECOS is the use of a consistent set of concepts and vocabulary for describing competences, as shown in Table 8-1. ECOS utilizes this comprehensive list for capturing competences and publishes it as an ECOS-card in RDF/XML format. This section presents how the ECOS ontology captures, presents and publishes competence in a formalized structure. In order to better understand this process, an example enterprise profile is considered i.e.:

> ABCJ Limited is a SME located in Leicester, UK. The company provides network consultancy and support to the academic and health sector. They offer a total IT support from managing your computers and networks and for writing software for them. ABCJ serves charity, education/Training, health/medical, ICT and Public Sectors. The key processes of the company are: IT networks, wireless networking, data applications etc. The company has skills in software development, IT networks and IT consultancy. ABCJ has close relationships and certifications from Microsoft, Novell, Cisco and many other industry standard technology companies. ABCJ always looks for project in the field of support and maintenance in IT networks.

In order to convert the above text profile into a semantic web resource, a form (ECOS-form) is provided to the user for filling it with the required information and standard codes with them. This form will generate an ECOS-card with the filled information for that enterprise. The detail about the process is as follows:

1. *Form fill up*: The ECOS ontology has four complementary concepts for capturing the key competences of an enterprise. An ECOS-form provides different fields, per se

Table 8-1: ECOS Ontology: List of Classes and Properties

Concept		Properties	Ontologies/Standards	Restrictions
General Information				
	Company Name	*ecos*:name | *ecos*:uri		
	Address	*ecos*:address | *ecos*:code	vCard (address) | NUTS (code)	
	Summary	xsd:string		
	Contact Person	*ecos*:person | *ecos*:code	vCard (person) | ISCO (code)	
	Key Person	*ecos*:person | *ecos*:code	vCard (person) | ISCO (code)	
Business Detail				
	Sector	*ecos*:code	SectorCode (code)	
	Product/Service	*ecos*:name | *ecos*:code | *ecos*:detail	CPC/ISIC (code)	
	Customer	*ecos*:Enterprise	*ecos* (Enterprise)	
	Preference	*ecos*:type | *ecos*:code | *ecos*:weight	CPV/ SectorCode/NUTS (code)	type={sector, region, process, product}
	Financial	*ecos*:name | *ecos*:value | *ecos*:time	iCalendar (time)	
Specific Information				
	Resource	*ecos*:name | *ecos*:detail | *ecos*:code | *ecos*:type	CPC/ISIC (code)	type={material, non-material}
	Process	*ecos*:name | *ecos*:detail | *ecos*:code	ISIC (code)	
	Unit	*ecos*:name | *ecos*:detail | *ecos*:code | *ecos*:responsibility | *ecos*:actor	ISIC (code)	
	Skill	*ecos*:name | *ecos*:detail | *ecos*:code	ISIC (code)	
	Plan	*ecos*:name | *ecos*:detail | *ecos*:code	ISIC (code)	
Business Record				
	Relation	*ecos*:type | *ecos*:Enterprise	*ecos* (Enterprise)	type={investor, supplier, customer, competitor, collaborator}
	Achievement	*ecos*:name | *ecos*:detail | *ecos*:code | *ecos*:type	CPC/ISIC (code)	type={patent, standard, publication}
	Past Project	*ecos*:name | *ecos*:detail | *ecos*:code | *ecos*:time | *ecos*:partner	*ecos* (partner) | iCalendar (time) | ISIC/CPC (code)	

ECOS concepts that need information from the user. The information required from users are under following categories:

- *General Information*: captures basic information about the company i.e. name, short summary, address, contact person and key person.

- *Business Details*: includes information about sector, product/services, customers, preferences and financial details of the company.

- *Specific Information*: captures tangible and intangible assets of the company, i.e. resources, skills, processes, organizational units and plans.

- *Business Records*: list information about past activities, relations and achievements of the company.

A screen shot of the ECOS-form filled with the information from ABCJ limited is shown in Figure 8-1. Once this basic information is filled, the next step is to fill appropriate codes for the required field.

2. *Codes Identification*: One of the objectives of proposing ECOS was to represent competences with a consistent set of vocabulary, widely accepted and shared among enterprises. In an attempt to articulate a comprehensive vocabulary, classifications defined by United Nations Statistics Division and European Union are considered in this research (**Appendix B**). These international classifications, including CPC, ISIC, NUTS and ISCO, are converted into semantic web resources using SKOS (Simple Knowledge Organization Scheme (Miles, et al. 2005)) and appended with the ECOS-form. Simple Knowledge Organisation Systems (SKOS) is a family of formal languages designed for representation of thesauri, classification schemes, taxonomies or any other type of structured controlled vocabulary. SKOS is built upon RDF and RDFS, and its main objective is to enable easy publication of controlled structured vocabularies for the Semantic Web. The SKOS model for standard classification can be obtained on the web at:

Figure 8-1: ECOS-form filled with basic information

Web Resource

```
8- Business and production services
   83- Other professional, technical and business services
      831 - Management consulting and management services; information technology services
         8313 - Information technology (IT) consulting and support services
            83132 - IT support services
```

Semantic Web Resource

```
<rdf:Description rdf:about="http://kmm.lboro.ac.uk/classification/cpc#83132">
  <skos:broadMatch rdf:resource="http://kmm.lboro.ac.uk/classification/cpc#8313"/>
  <skos:hiddenLabel>83132</skos:hiddenLabel>
  <skos:prefLabel>IT support services</skos:prefLabel>
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
</rdf:Description>
```

Figure 8-2: Classification Code as Semantic Web Resource

- *Sector code*: http://kmm.lboro.ac.uk/classification/sec

- *NUTS code*: http://kmm.lboro.ac.uk/classification/nuts

- *ISIC code*: http://kmm.lboro.ac.uk/classification/isco

- *CPC code*: http://kmm.lboro.ac.uk/classification/cpc

- *ISCO code*: http://kmm.lboro.ac.uk/classification/isco

- *CPV Code:* http://kmm.lboro.ac.uk/classification/cpv

A simple example, explaining how a CPC code available as a web resource is represented as a semantic web resource, is shown in Figure 8-2. These classifications are used in the ECOS model to provide a standard vocabulary for competence representation. A screen shot of an ECOS-form with blanks for classification codes is shown in figure 8-3. From Figure 8-3, it is clear that a user has to select an appropriate code from the list (CPC codes is shown in Figure 8-3) and fill the required blanks manually in the form. The concepts available in the ECOS-form that require classification codes are as follows:

- *Sector code*: Sector, Customer, Preference and Relation

- *NUTS code*: Address

- *ISIC code*: Process, Resource, Unit, Plan, Past Project and Achievement

- *CPC code*: Products/Services, Resource, Past Project and Achievement

138

Figure 8-3: ECOS-form with blanks for Classification Codes

- *ISCO Code*: Contact Person, Key Person
- *CPV Codes*: Preference

The complete information available in the company's profile can be entered in the ECOS form and appropriate code can be selected using the provided lists. In this research, the process of filling the information in the ECOS-form is carried out manually. Although, this process can be performed automatically, this is beyond the scope of this research.

Figure 8-4: ECOS-card in two formats, HTML/XML and RDF/XML.
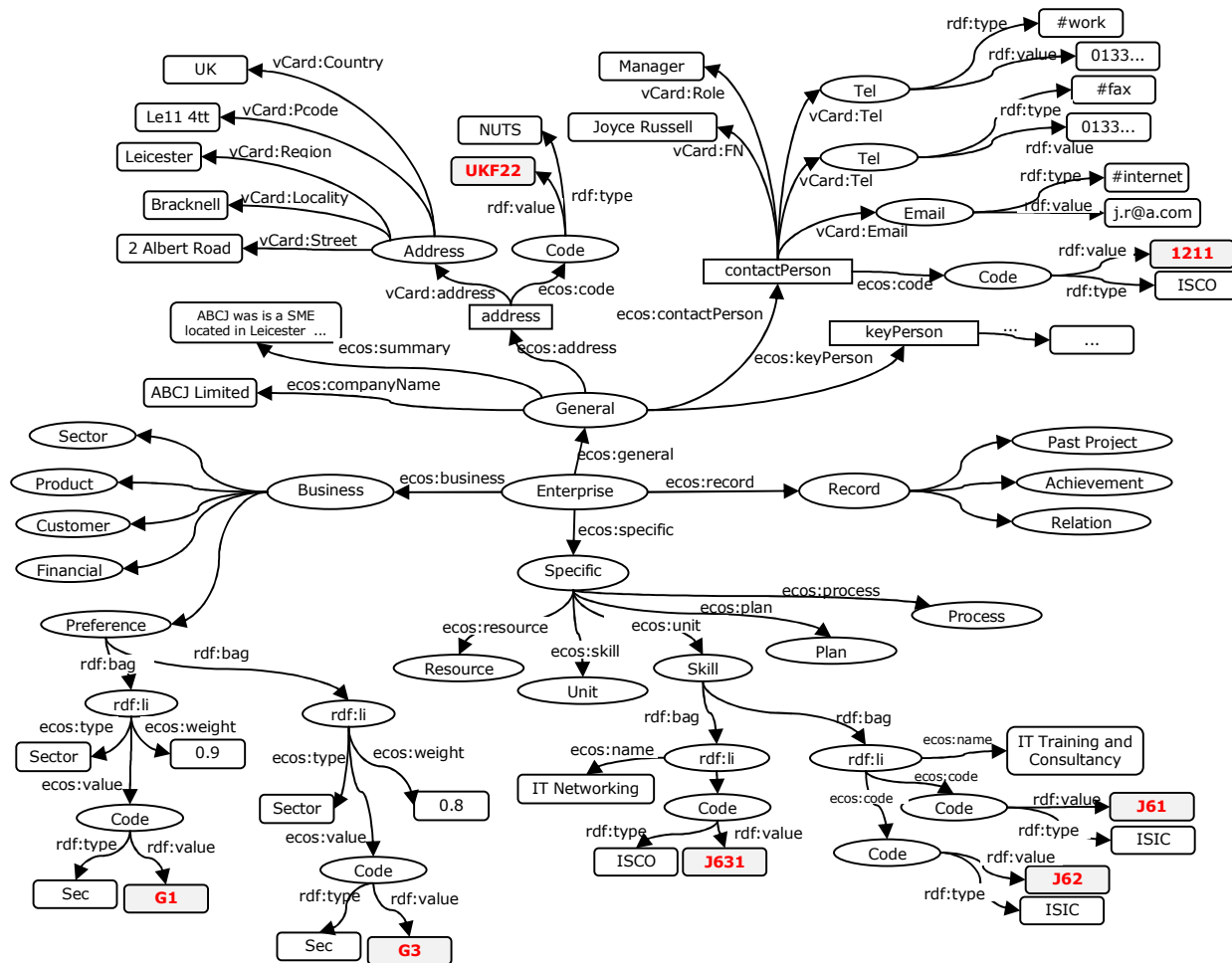
Figure 8-5: Tree Representation of ECOS-card generated in Figure 8-4

They offer a total IT support from managing your computers and networks and for writing software for them.

HTML/XML

Name: IT Support
Code: cpc#83131, cpc#83132
Detail: ABCJ provides full IT support.

RDF/XML

```
<ecos:Product>
<ecos:name>IT Support</ecos:name>
<ecos:code rdf:parseType="Resource">
  <rdf:value>cpc#83131</rdf:value>
  <rdf:type rdf:resource= "http://kmm.lboro.ac.uk/classification/cpc#83131" />
</ecos:code>
<ecos:code rdf:parseType="Resource">
  <rdf:value>cpc#83132</rdf:value>
  <rdf:type rdf:resource="http://kmm.lboro.ac.uk/classification/cpc#83132"/>
</ecos:code>
<ecos:detail>ABCJ provides full IT support</ecos:detail>
</ecos:Product>
```
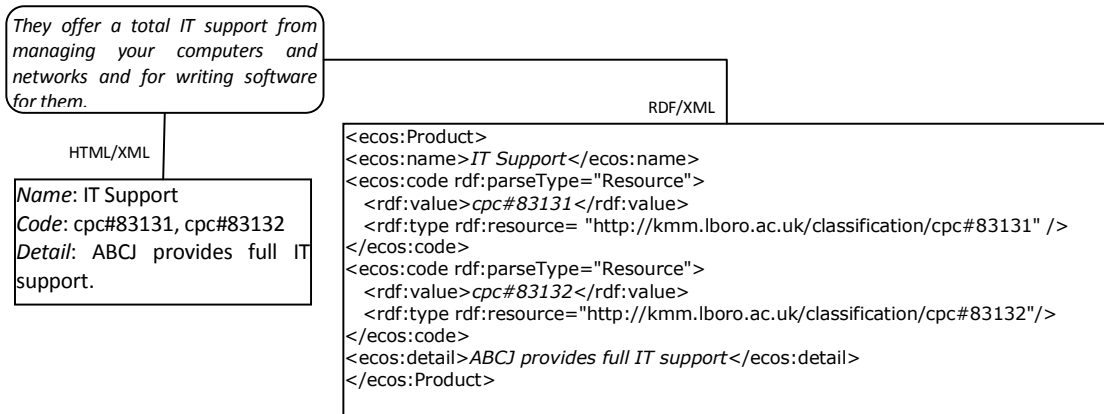
Figure 8-6: Storing Competence in two formats

ECOS-card Generation: Once the enterprise information and appropriate codes are filled in the required fields, an ECOS-card can be generated from the ECOS-form. The ECOS-card generated for the above example is shown in Figure 8-4. A tree representation of the ECOS-card generated from the ECOS-form is shown in Figure 8-5. For brevity, the tree structure of the ECOS-card is not shown completely. Only 3 concepts namely, ECOS:companyName, ECOS:Address, ECOS:contactPerson, ECOS:Skill and ECOS:Preference are shown in figure 8-5. The remainder of the concepts are self explanatory.

As mentioned earlier, the ECOS-card for ABCJ limited is provided in two formats, i.e. HTML/XML (human readable) and RDF/XML (machine readable), to describe the company's basic information and competences labelled by predefined classification codes to it. Figure 8-6 shows how the same information is represented in two different formats. The main advantages of this card are: (1) a computer can view this information as an explicit tree structure (as shown in Figure 8-5), (2) it can store information in any decentralized location, i.e. in different types of databases (essentially triple store) and in different machines, (3) it can find information and accomplish tasks without human direction (4) computers can extract information from this card using the SPARQL query language, as discussed in Chapter 4.3.4.
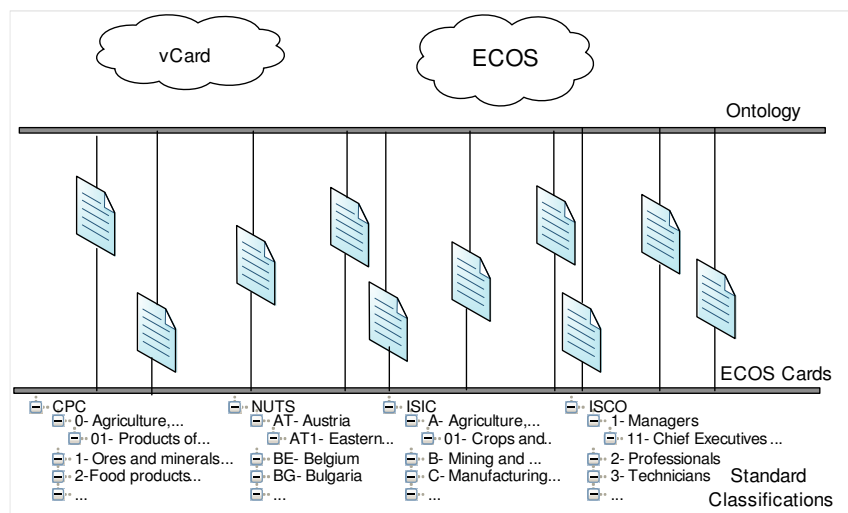
Figure 8-7: Links between ECOS-card, Classifications and Ontologies

This approach was repeated to create ECOS-cards for 25 enterprises. As evident from prior discussions and the above demonstration, an ECOS-card presents enterprise competences by using a consistent set of concepts and vocabulary and can then publish/store this information as a semantic web resource. These semantic profiles are
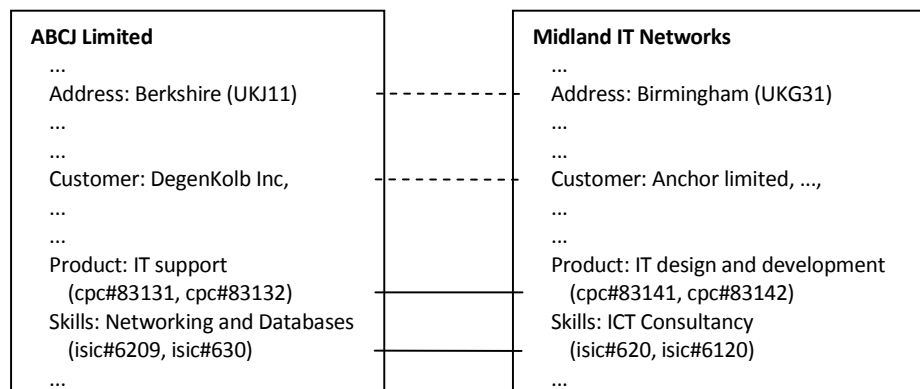
Figure 8-8: Types of relations between 2 ECOS-cards

directly linked with standard classifications and existing ontologies, as shown in Figure 8-7, and are thus indirectly linked with each others. These hidden relationships can be used for matching ECOS profiles and identifying relationships between them. For example, in

Figure 8-8, the two enterprises, ABCJ Limited and Midland IT Networks, are in different cities and working with different customers. But from their ECOS-cards, it is evident that they have similar products and skills due to similar CPC and ISIC codes, respectively. Such relations can be used as a basis for building social relationships and supporting various enterprise applications, network analysis, project selection, etc.

In the next section, these hidden relations are identified and analyzed for proposing potential collaborations.

8.4. Creation Stage

An example collaboration pool with 25 ECOS-cards has been analyzed and compared at this stage of the demonstration of the virtual enterprise life cycle. This analysis is carried out in three steps, as discussed in Chapter 5. Initially, the ECOS-web module is used to identify direct and hidden relations among enterprise profiles. In the ECOS-track module, tender documents are selected from the existing tender-alert services (Sell2Wales) and evaluated against ECOS-cards. The ECOS-match module assesses the suitability of pool members against the tender requirements and proposes potential virtual enterprises. Finally, a list of corresponding enterprises with tenders and potential partners is generated.

8.4.1.   *ECOS-web: Maintaining Collaboration Pool*

ECOS-web provides a visual and mathematical analysis of enterprise relations, ranging from general links (e.g. same customers, project partners, etc) to hidden relationships (e.g. similar process, product, location etc). As a demonstration for ECOS-web, initially a small and simple example with 6 enterprises from collaboration pool, each with 2 properties is considered. Table 8-2 shows the list of enterprises with two properties, their customers and address (only city name and NUTS code is considered here). The ECOS-web graphs generated for representing direct and hidden relations among ECOS profiles are shown in Figure 8-9 and Figure 8-10 respectively. The details about these graphs are as follows:

In table 8-2, ABCJ Limited (e-1) is customer of Anchor Limited (e-3) and BCRA Company (e-4). These direct relationships are shown as solid lines in Figure 8-9.

In Figure 8-10 dashed lines are used for representing hidden relations among enterprises. These hidden relations are identified based on the NUTS code assigned to their addresses. The weights assigned to these hidden relationships are evaluated by counting the number of links (i.e. number of nodes + 1) lying between two nodes in NUTS classification hierarchy (as discussed in Section 5.2.3).

For example, in Figure 8-11, the number of nodes between two NUTS codes assigned for the two cities Leicester (UKF21) and Derby (UKF11) are three (i.e. D and N; East Midlands; L, R and NS). Therefore, the edge weight for the link between Leicester and Derby will be 0.25.

Table 8-2: An Example for ECOS-web demonstration

No.	Name	Address	NUTS	Customer
e-1	ABCJ Limited	Bracknell, Berkshire	UKJ11	e-10
e-2	Alan M. Wye and Co	Street, Luton, Bedfordshire,	UKH21	e-11
e-3	Anchor Limited	Loughborough, Leicester	UKF22	e-1
e-4	BCRA Company	Torwood Close, Middlesex	UKI22	e-1
e-5	Berger Firm	Smethwick, West Minister, London	UKI11	e-2
e-6	Cary Cam Engineering	Transport Ave., Brentford, Middlesex	UKI23	e-3, e-4

Similarly, all the edge weights are calculated for relations shown in Figure 8-10.

The above two graphs (Figure 8-9 and 8-10) represent relations between enterprises based on individual properties of enterprises. In order to evaluate the net importance of the enterprises, a single graph is designed by adding the edge weights of different ECOS properties along with the preferences of users for different properties (similar preference is considered). For example, the edge weights between Cary Cam and BCRA is 1 for customer relationship and 0.5 for address, therefore net weight in combined graph will be 0.75. The derived graph representing the total weight between different enterprises is shown in Figure 8-12. It is evident from Figure 8-12 that the derived graph consists of various edges with low weight, thus making the entire graph cumbersome. Therefore, a simple graph with edge weight more than 0.10 is also derived and shown in Figure 8-13.

The derived graphs, shown in Figure 8-12 and Figure 8-13, can be viewed as social network graph of enterprises that maps all the relevant ties between all the enterprises in the collaboration pool. In many ways, network graphs are the obvious solution for trying to visually represent a network structure. No other graphical system functions as much as a proficient method for representing connections between nodes. Sociologists have been studying the social network graphs for decades. The most widely known study in this area is the *Six Degrees of Separation* (Watts 2003), which came with an anecdotal and scientific observation that everyone on this planet is connected to other –
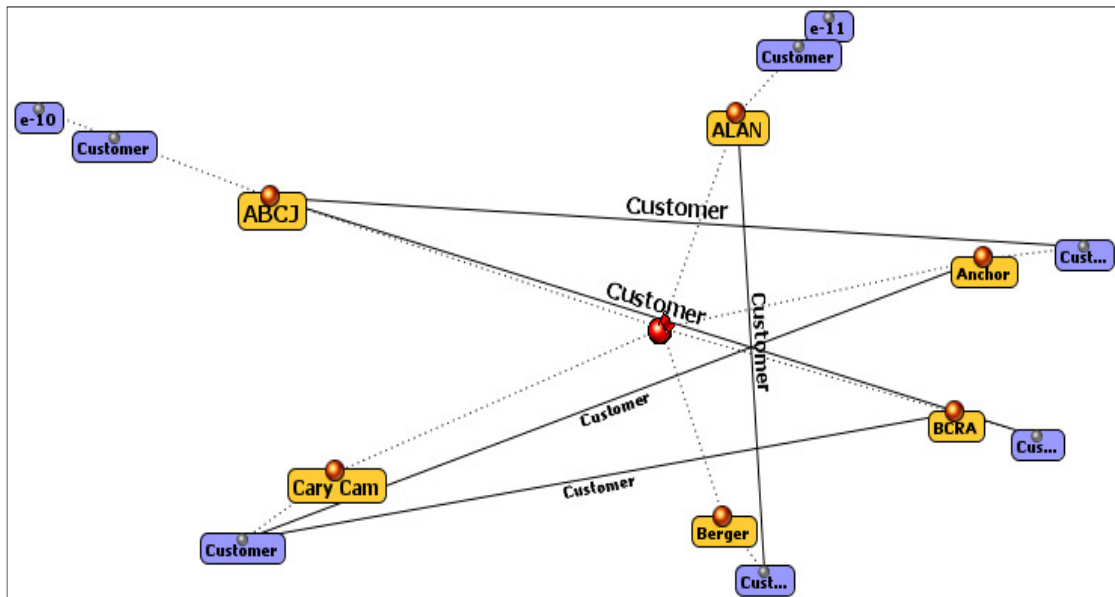
Figure 8-9: ECOS:Customer, Direct Links in ECOS-web graph
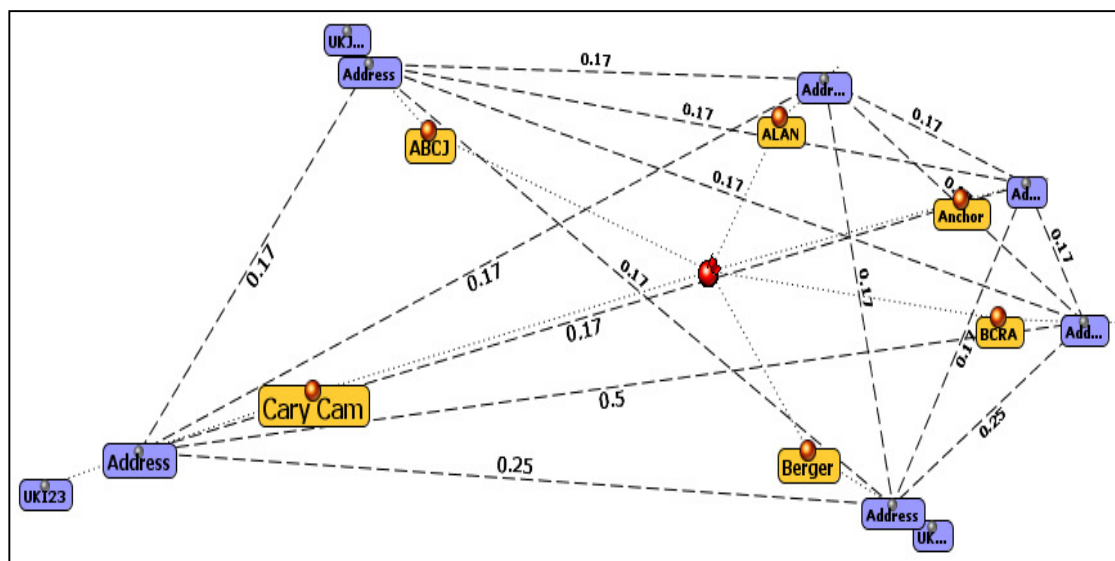
Figure 8-10: ECOS:Address, Hidden Links in ECOS-web graph
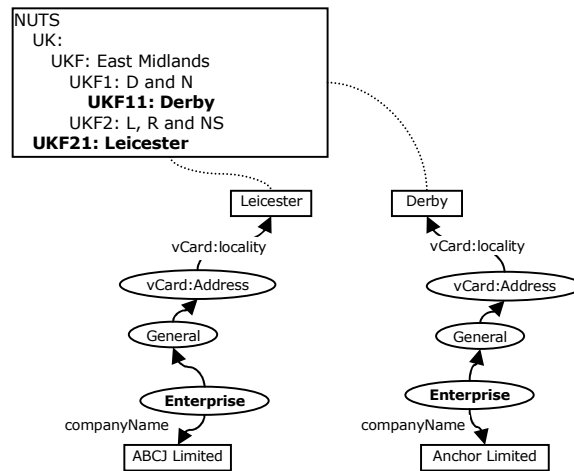
Figure 8-11: An example for Distance Calculation
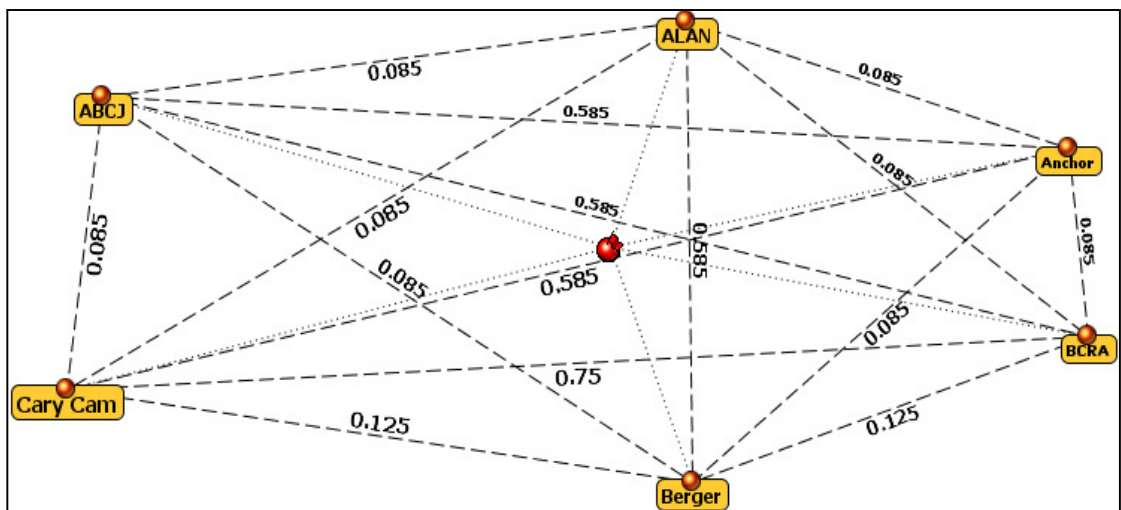
Figure 8-12: ECOS-web graph with Average weights
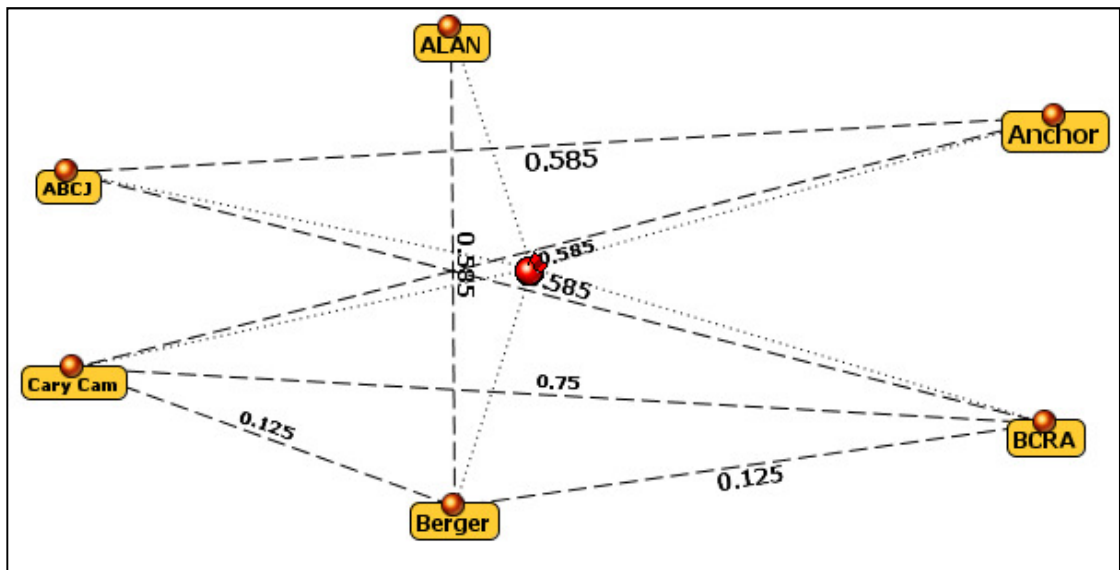
Figure 8-13: ECOS-web graphs with higher edge weights (Weights> 0.10)

Table 8-3: Centrality measure for Figure 8-12

Enterprise	Degree	Betweenness	Closeness
ABCJ Limited	1.425	0	0.135
Alan M. Wye and Co	0.925	0	0.085
Anchor Limited	1.425	0	0.135
BCRA Company	1.63	3	0.157
Berger Firm	1.005	4	0.1
Cary Cam Engineering	1.63	3	0.157

Table 8-4: Centrality measure for Figure 8-13

Enterprise	Degree	Betweenness	Closeness
ABCJ Limited	1.17	0	0.076
Alan M. Wye and Co	0.585	0	0.048
Anchor Limited	1.17	0	0.076
BCRA Company	1.46	3	0.088
Berger Firm	0.835	4	0.056
Cary Cam Engineering	1.46	3	0.088

no more than six people away. Similarly, organizations have also been using social networks to examine how they are related with others, characterizing the informal connections that link companies together, as well as associations and connections between individual employees at different organizations (Ahuja, Carley 1999). The ECOS-web graph developed in this research can also play a key role in organizations to identify their direct and indirect relations with other companies in the collaboration pool.

ECOS-web graphs provide an alternate view for visualizing and analyzing relations, where attributes of enterprises are less important than their ties and links with other enterprises in the pool. The study of attributes reveals information that is inherently contained within and not between enterprises. Analysts sort enterprises based on their attributes and determine which outcomes are disproportionately common to enterprises with particular attributes. In contrast, the study of relationships embedded in network structures can help analysts to explain macro-level patterns, e.g. a large number of enterprises are closely related not because they are similar, but as a large number of enterprises acting on one another to shape one another's outcome. For example, analysts might find it difficult to explain certain relations such as, *Berger Firm* is the most important company in the pool as it is the only firm strongly linked with *Alan M. Wye*. Similarly, despite BCRA and Cary Cam having different customers, the strongest link in the network is between them.

The network graph is further used in this research for evaluating the importance of enterprises and their pattern of relationships in a network. The measure considered in ECOS-web module for network analysis is centrality of the network (discussed in Section 5.2.3). The three measures (i.e. degree, closeness and betweenness) discussed for evaluating the centrality of the network structure are calculated for two derived graphs (with and without weights < 0.10, respectively) and results are provided in Table 8-3 and Table 8-4. The significance of these three measures is as follows (mainly Table 8-4 is used here, as it contains links with significant weights):

- The degree centrality shows direct connection of an enterprise with other enterprises. Enterprises with high degree show strong ties with others and thus higher opportunities for developing relations with others. From Figure 8-13 and Table 8-4, although Berger, BCRA and Cary Cam are linked with 3 other companies but last two are more closely related to others and have strong customer relations in total.

- Betweenness identifies a node's position within a network in terms of its ability to make connections to other pairs or groups in a network. Enterprises with higher betweenness lie more between enterprises and play an important role in the network with a potential for control over others. In the above example, Berger has the highest betweenness as it is the only link with ALAN in the network.

- Closeness centrality measures the shortest distance of an enterprise between all other nodes in the graph. Thus, a high closeness value shows that the enterprise has close relations with others in the network. In the above example, ALAN has the lowest closeness centrality as it is not directly related with any other enterprises. It can be noticed that, although Berger has a high betweenness centrality but a low closeness centrality as it is not directly related to ABCJ and Anchor.

Similarly, the ECOS-web graph can be generated for bigger collaboration pools and analysis can be carried on the generated network graphs. In order to create the ECOS-web graph with selected ECOS properties, a web application is developed in this research. This web application provides a list of ECOS properties and ECOS cards available in the collaboration pool for generating the ECOS-web graph. A snippet of the web application is shown in Figure 8-14. The ECOS-web graph generated for 25 enterprises present in the collaboration pool with three properties, i.e. address, products/services and skills, is shown in Figure 8-15 and Figure 8-16. The Figure 8-16 is a derived graph representing relations between enterprises with edge weight greater than 0.30. These edge weights are calculated for each enterprise's relations and are further used for
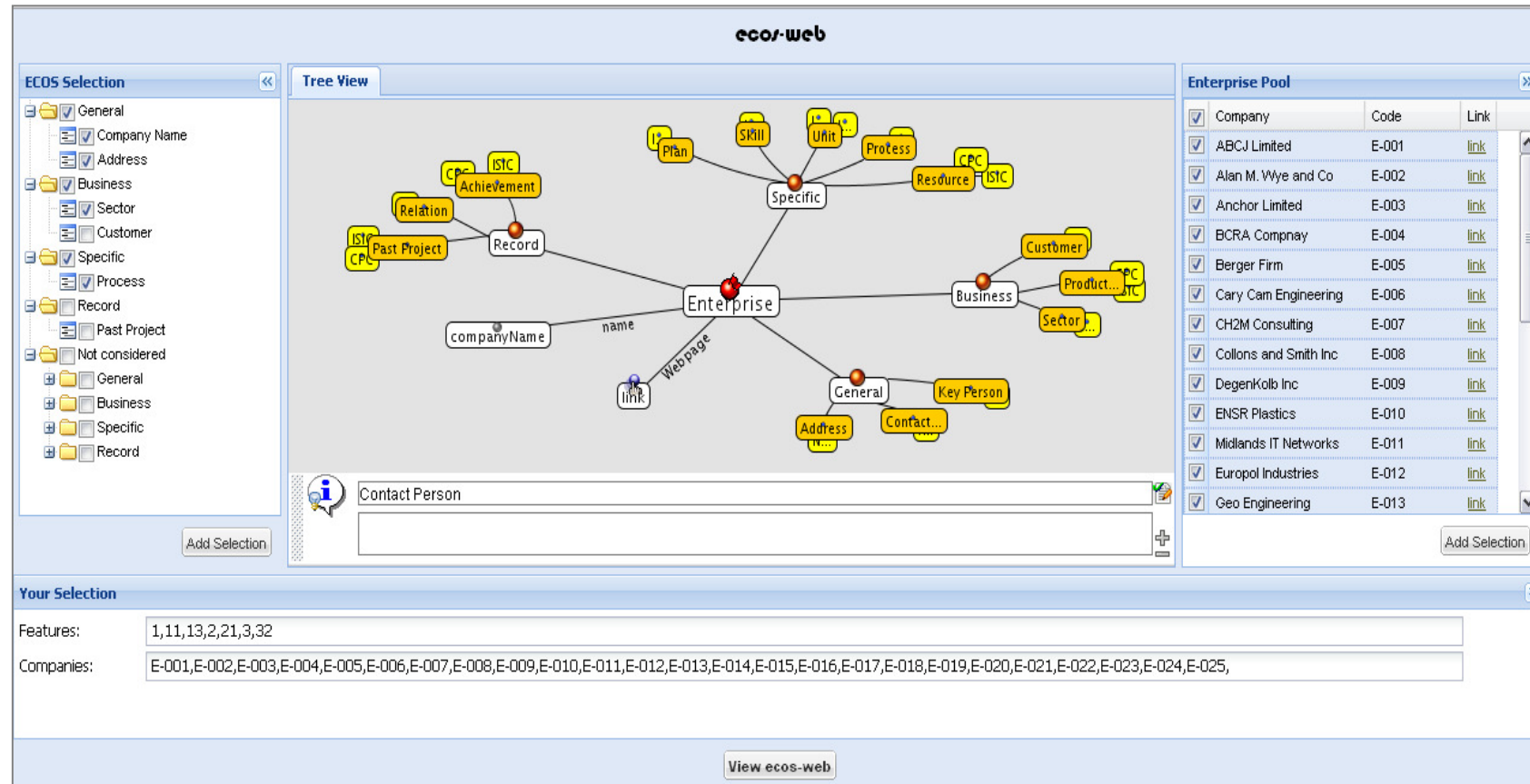
Figure 8-14: Snippet of ECOS-web Application
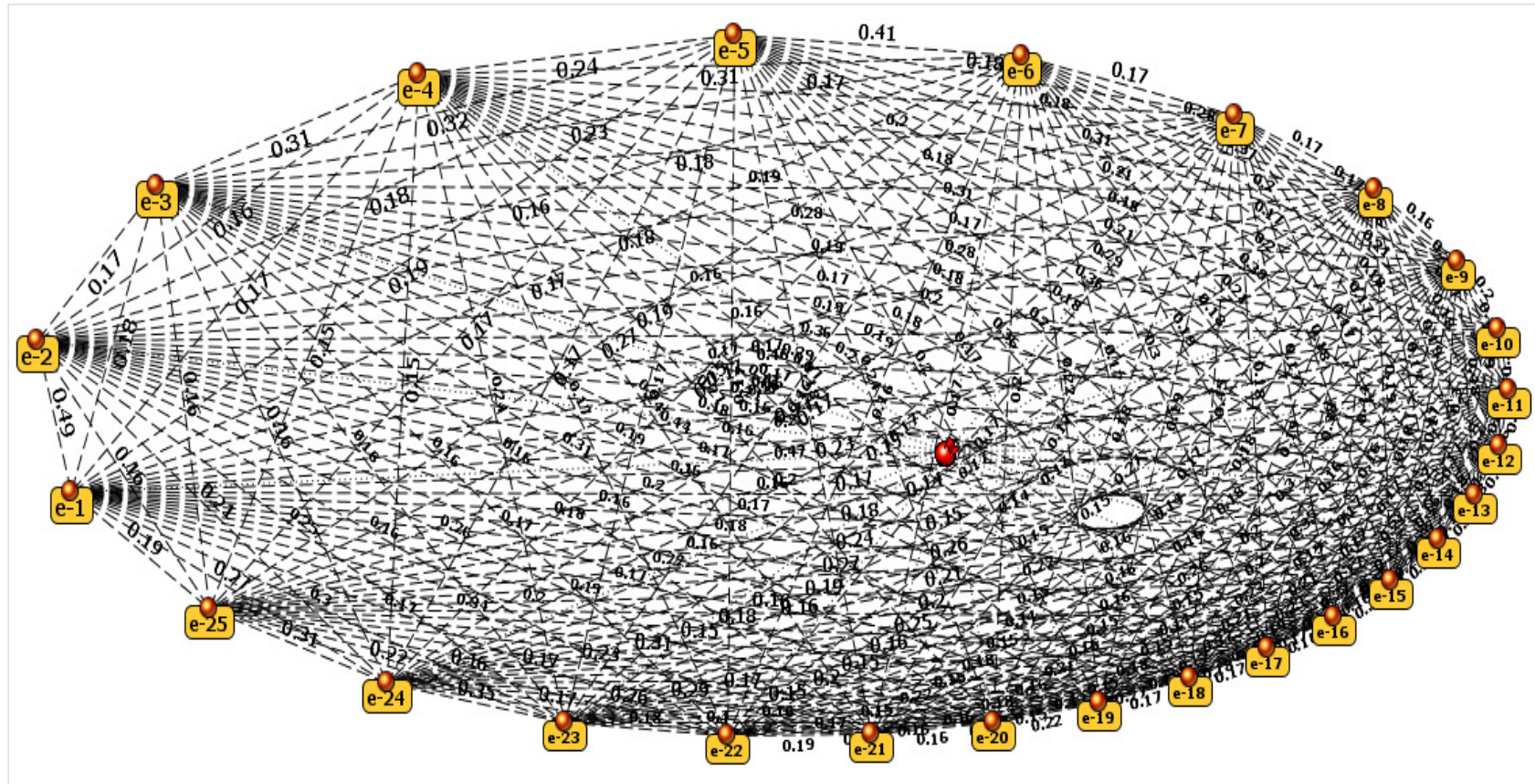
Figure 8-15: ECOS-web graph for 25 Enterprises for average weights of ecos:product, ecos:skill and ecos:address
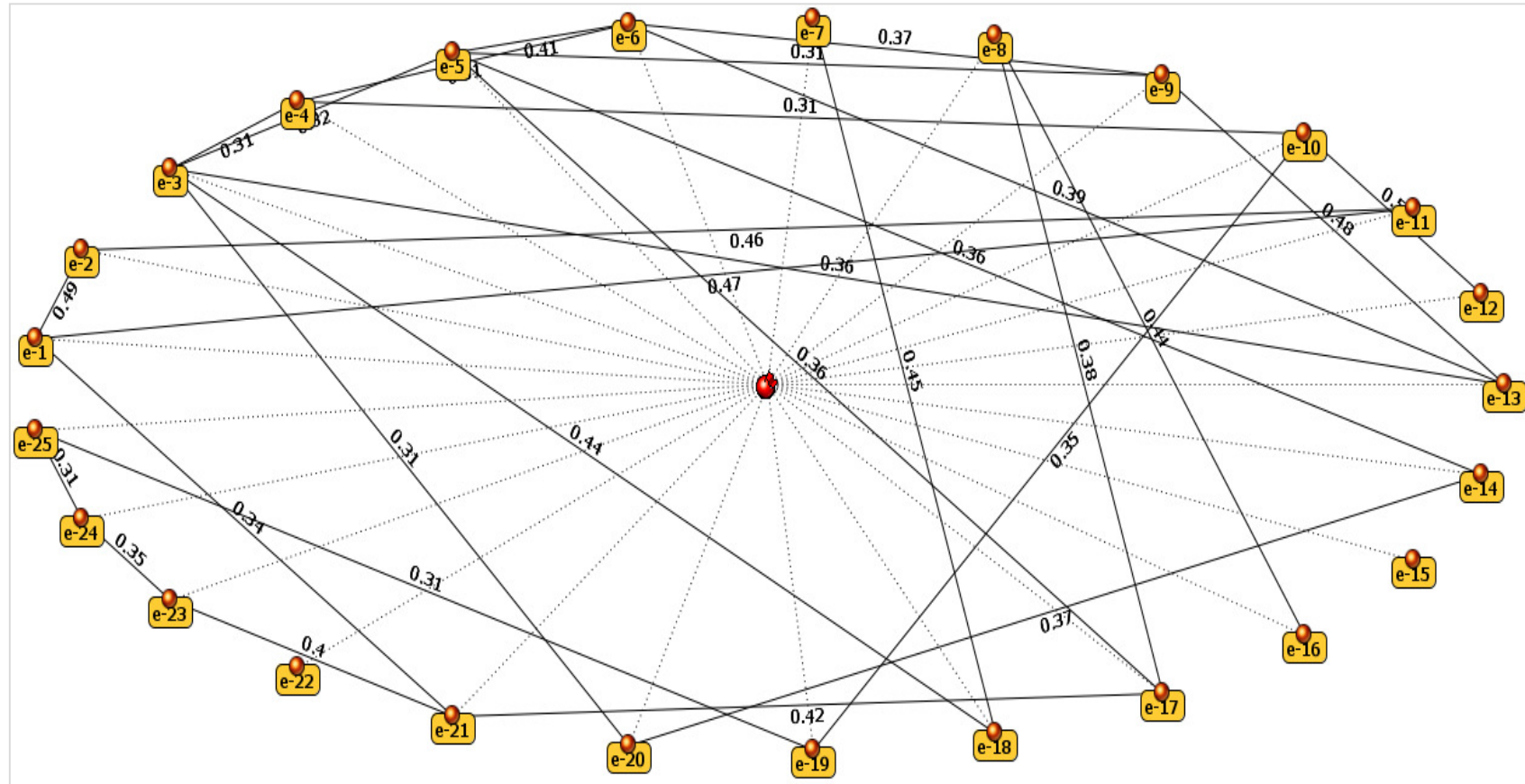
Figure 8-16: ECOS-web graph with edge weight> 0.30

calculating the centrality measures. The results evaluated for the three centrality measures i.e. degree, closeness and betweenness are provided in **Appendix F**.

The ECOS-web graphs generated in this research and shown as screenshots are created using Treebolic Generator, a Java application developed for representing hierarchical data in tree structure (Details in **Appendix E**).

8.4.2.   *ECOS-track: Prioritizing Business Opportunities*

The task of the ECOS-track module is to compare ECOS profiles against the available business opportunities. In this case study the business opportunities are taken from a Tender Alert Service, Sell2Wales. These notifications are based on the profile information provided by the enterprise subscribed to the tender notification site. In this research, the entire collaboration pool is registered with this tender-alert service as a single entity and tender notifications are received from it. Subsequently, the relevant information is extracted from the obtained notifications and weighted against the capabilities of enterprises present in collaboration pool.

However, the tender notices obtained from *Sell2Wales* are natural language documents such as html page, PDF file, word document, etc. The first task of the ECOS-track module is to convert these web documents into a unified format. The module gleans information items obtained in different formats and transforms them into an unstructured text document i.e. .txt file. Next, it extracts relevant information from the .txt files. The contract notices from Sell2Wales are usually drafted in a standard format given by Official Journal of the European Union (OJEU) (SIMAP). In these contract notices, the relevant information material is separated into different sections that provide a clear idea about the tender. The task of the module is to extract relevant information from these sections. The information item essential for the tender selection are the tender requirements (given in form of CPV codes) and location of the tender. Although, other information is also gathered from the notice, the CPV codes are key data required for tender-enterprise comparison.

Table 8-5: Calculation of Utility Factor

		Tender Document (CPV Codes with Rating >0)						
		72000000	72530000	72531000	72532000	72315000	72315100	72318000
Enterprise (CPV Codes)	72000000	**1**	0	0	0	0	0	0
	72300000	0	0	0	0	0.25	0.25	0.25
	72315000	0	0	0	0	**1**	0.75	0.5
	72315100	0	0	0	0	0.75	**1**	0.5
	72315200	0	0	0	0	0.75	0.75	0.5
	72318000	0	0	0	0	0.5	0.5	0.5
	72530000	0	**1**	**0.5**	0.5	0	0	1
	72532000	0	0.5	0.5	**1**	0.25	0.25	0
	Maximum	**1**	**1**	**0.5**	**1**	**1**	**1**	**1**

Sum	6.5
Tender Rating =(Sum/18, Number of CPV codes)	0.361
Location	1
Net Utility Factor=(0.8*Tender Rating + 0.2*Location)	0.489

A sample tender notice considered here is for establishing the web-hosting services for Scotland Government (Sell2Wales). The tender document is available at the Sell2Wales Website[7]. The list of CPV codes extracted from this document is: 72000000, 32551400, 32412120, 32410000, 32412000, 32412100, 32412110, 72530000, 72531000, 72532000, 72315000, 72315100, 72318000, 50334110, 64210000, 64212000, 64214000, and 32430000. The location of the project provided in the tender is Leicestershire and elsewhere in the UK.

The CPV codes extracted from this sample tender document is compared with the preferences of the enterprises provided in ECOS profiles. The individual rating is calculated for each CPV code in the tender document against the CPV codes provided by each enterprise. For example,

---

[7] https://www.sell2wales.co.uk/notices/displayOJ.html?NoticeId=170291&activity=Mzc4Mjc%3d

Table 8-6: Utility factor for Enterprises against Tender Opportunity

Company	Name	Preference (0.8)	Location (0.2)	Average
E-1	ABCJ Limited	0.389	1	0.511
E-2	Alan M. Wye Co.	0.361	1	0.489
E-3	Anchor Limited	0.389	1	0.511
E-7	CH2M Consulting	0.528	1	0.622
E-11	Midlands IT Networks	1	1	1
E-14	Golder Technology	0.569	1	0.656
E-20	Pulp Inc.	0.444	1	0.556
E-21	Seatle National Groups	0.917	0	0.733
E-22	Shardal Ltd	0.556	0	0.444
E-23	SIMS Ltd	0.431	0	0.344

The list of CPV codes provided by Alan M Wye Co. as their preferences is: 72000000, 72300000, 72315000, 72315100, 72315200, 72318000 and 72532000. On comparing this list of CPV codes from Alan M Wye Co. against the list in tender document, as shown in Table 8-5, the tender rating evaluated for the enterprise is 0.361.

On the other hand, as Alan M Wye Co. is located in Bedfordshire, UK (shown in Table 8-2) and the location requirement in the tender document is anywhere in UK, the rating for location assigned to the enterprise is 1.

Finally, the two rating are combined as per the preferences of the users and net utility factor is evaluated for the enterprise. In this example, the preference weights for tender and location is 0.8 and 0.2 respectively, and therefore the net utility factor calculated for Alan M. Wye Co. is 0.489.

Similarly, the total utility factor for the tender document for each enterprise is evaluated, and the results are provided in Table 8-6. From table 8-6, it is evident that enterprise E-14 with utility factor 1 is able to satisfy all the requirements of tender document. However, there are various other enterprises that are able to provide input in the project but cannot carry out the whole tender individually. The ECOS-match module is used here to analyze different combinations of enterprises and suggest potential collaborations.

8.4.3.   *ECOS-match: Suggesting Enterprise Collaborations*

The task of the ECOS-match module is to list the potential collaborations among enterprises based on utility factors evaluated for each tender document and relations

Table 8-7: Calculation of Objective value for a Virtual Enterprise

		E-2	E-7	E-20	Net Weight
Tender Document (CPV Codes)	72000000	1	1	0	0.8
	32551400	0	0	1	1
	32412120	0	0	1	1
	32410000	0	0	1	1
	32412000	0	0	1	1
	32412100	0	0	1	1
	32412110	0	0	1	1
	72530000	1	1	0	0.8
	72531000	0.5	0.5	0	0.3
	72532000	1	1	0	0.8
	72315000	1	1	0	0.8
	72315100	1	1	0	0.8
	72318000	1	1	0	0.8
	50334110	0	0	1	1
	64210000	0	1	0	1
	64212000	0	1	0	1
	64214000	0	1	0	1
	32430000	0	0	1	1
				Sum	16.1
				Tender Rating	0.894444
				Location	1
				Utility Factor	0.915556

identified in the ECOS-web module. The relations identified in the ECOS-web module are incorporated to advance the partner selection process by including direct relations and hidden relations among enterprises, e.g. similar competences, past projects, etc. Therefore, different combinations of enterprises are analyzed from the collaboration pool and weighted against given tender requirements to propose potential virtual collaborations. The three types of weights considered in this research are, 1) Net utility factor calculated in ECOS-track module, 2) Average of the edge weights assigned for the enterprise relations in ECOS-web module, 3) Combination of above two weights using

Table 8-8: Potential Virtual Enterprises for Tender Opportunity

Virtual Enterprise	ECOS-track (0.5)	ECOS-web (0.5)	Average
E-11	1		
E-2, E-20	0.978	0.588	0.783
E-1, E-14	0.967	0.585	0.776
E-2, E-14	0.944	0.581	0.763
E-1, E-11	0.938	0.735	0.837
E-3, E-11	0.938	0.587	0.762
E-1, E-7, E-20	0.927	0.457	0.692
E-1, E-3, E-14	0.921	0.476	0.699
E-2, E-7, E-20	0.916	0.449	0.682
E-3, E-7, E-20	0.916	0.481	0.698
E-1, E-14, E-20	0.912	0.496	0.704
E-1, E-2, E-7, E-20	0.892	0.423	0.658
E-1, E-3, E-14, E-20	0.887	0.44	0.663
E-1, E-3, E-7, E-20	0.864	0.404	0.634
E-2, E-3, E-14, E-20	0.864	0.434	0.649
E-2, E-3, E-7, E-20	0.853	0.398	0.625
E-1, E-2, E-3, E-7, E-20	0.83	0.378	0.604
E-1, E-2, E-3, E-14, E-20	0.813	0.401	0.607
E-1, E-3, E-14, E-20, E-23	0.81	0.383	0.596
E-1, E-2, E-3, E-7, E-14, E-20	0.752	0.355	0.554
E-1, E-2, E-3, E-7, E-20, E-23	0.751	0.348	0.549

the user preferences. A simple example explaining the process of calculating weights for a combination of enterprise is as follows:

Consider a combination of three enterprises (E-2, E-7, E-20) from the collaboration pool shown in Table 8-6. Only 10 enterprises are considered in the collaboration pool, because other enterprises do not satisfy any requirement from the tender document.

Initially, the virtual enterprise is evaluated against the tender document to calculate the tender utility factor for this combination. Assuming this combination as a single enterprise, at first the tender rating and location rating is calculated and then combined for getting the net utility factor for the virtual enterprise (as calculated in previous section). However, in a virtual enterprise two or more enterprises can satisfy the same tender requirement and thus increase the factor of competition in the collaboration. In order to reduce this similarity, the enterprises satisfying the same tender requirements are reduced by a factor, i.e. (Total rating/Size of Collaboration Pool). The calculation of net utility factor (i.e. =0.916) for this combination of enterprises is shown in Table 8.-7.

Then, the edge weights evaluated for the enterprise relations in the ECOS-web module are combined to calculate the net weight for this combination. The edge weights for these enterprises are:

$$w = \begin{pmatrix} 1 & 0.169 & 0.163 \\ 0.183 & 1 & 0.179 \\ 0.172 & 0.172 & 1 \end{pmatrix}$$

The average of these weights is calculated to get the net weight for enterprise relations, i.e. =0.449.

Finally, the above two weights are combined to get the net objective value for this combination of enterprises, i.e. 0.682.

Similarly, all the combinations of enterprises in the collaboration pool are evaluated against the tender document. The potential virtual collaborations with significant tender ratings, average ECOS-web ratings and their average values are provided in Table 8-8. The above analysis can facilitate the process of selecting the most suitable tender and appropriate partners for virtual enterprises.

8.5. Operation Stage

Once an appropriate tender and the suitable partners are selected for virtual collaboration, the partners team up to work as one organization in the operation stage of the virtual enterprise. In this stage, partners collaborate and integrate their core competencies to satisfy the project goals defined in the tender requirements. At the start of operation phase, the focus is on defining the terms of engagement and collaboration agreement, identifying the functional requirements, and assigning tasks to each partner to achieve the goals of the collaboration. Subsequently, each partner starts working on their assigned tasks using their core competences and sharing information (as agreed) with other business partners. During the execution of a project, a virtual enterprise will inevitably face various unforeseen problems, such as inconsistency in information sharing and communication, conflicts in objects of interest, disagreement with partners etc.

It is not possible to comprehensively address all types of problems that a virtual collaboration is likely to face during its operation stage. Therefore, in this research the problem considered is the inconsistency in information sharing and communication. A specialized knowledge service is proposed to facilitate this process. Consider the following example to understand the scenario used to illustrate this problem and proposed solution:

> A virtual enterprise with 3 partners, 1, 2 and 3, from three different sectors, Information and Communication, Construction and Financial and insurance activities are working on a project. In this collaboration, each enterprise works on their specific tasks and share information with other two companies to achieve the final goal of virtual collaboration. However, information sharing and communication between these cross-disciplinary companies is often hindered by the lack of clarity in terms and vocabulary used. Generally, companies develop and use their own vocabulary and terms for particular issues, elements or activities. Hence, when these three multi-disciplinary enterprises are brought together, two types of problem in information sharing can occur, i.e. different terminologies are used to describe the same meaning or the same terminology is being associated with different meaning.

It is evident that lack of semantics about the meaning of contents in a particular domain can affect the overall meaning of the information and the way in which it is

Figure 8-17: TEXT2RDF Application (Screenshot)

interpreted. The TEXT2RDF framework is proposed to support this process by providing explicit semantic descriptions to the information resources and reducing the potential for errors caused by data duplication and the resulting inconsistencies. It extracts information from the web documents, converts it into a semantic document (RDF format), and hyperlinks the terms and keywords with domain specific meanings.

At first, the TEXT2RDF approach extracts a bag of phrases from a text document and coverts it into a semantic web resource. This process is carried out in 5 steps, as discussed previously in Section 6.4, i.e. 1) Converting a web document into a text document, 2) Parsing and Filtration of the text, 3) Analyzing the text to get bag of phrases (BOP), 4) Creating an RDF document from the text, and 5) Evaluate the accuracy of the results. A Java based web application is coded in this research to demonstrate the implementation of proposed TEXT2RDF approach. The techniques and technologies used to realize this 5-step process and develop the web application are as follows:

1. *Document Processing*: Initially, a web-based form is created, where users can insert the text data and get the content in an RDF format. A snapshot of the webpage is

shown in Figure 8-17. Although, the text documents in the enterprises are mainly in file formats (e.g. .doc, .pdf, webpage etc), this web page is developed to lucidly demonstrate the process and results of the approach.

2. *Parsing and Filtration*: A web-service is coded in the Java language to read the text data from the webpage, parse the text content and extract terms and phrases from it. At first, the text data is tagged with the corresponding part of speech (POS) using Stanford NLP parser (Klein, Manning 2003) and the terms and phrases are listed from the text content. Then, a list of stop words[8] is used to remove the irrelevant words present in the list. The implementation of this process on a simple example has been explained previously in Section 6.4.2.

3. *Text Analysis*: In this step, a bag of phrases (BOP) is generated by labelling the terms and phrases with appropriate labels, such as keyword, phrase, iPhrase, cardinal, etc. The list of rules discussed in Section 6.4.3 is used to identify the labels for the terms and phrases present in the text document.

4. *RDF Generation*: The software Jena, a semantic web API in Java, is used to convert the text content and BOP into a semantic web document, i.e. RDF format. The ontology for document-model, shown in Figure 6-1, is used as the RDF schema for creating the RDF documents.

5. *RDF Revision*: In this research, the accuracy of terms and phrases extracted from the text content is analyzed against the terms and phrases extracted manually from the text document.

The BOP is further annotated with domain specific meanings using a hyperlink approach. The terms and phrases extracted from the document are matched with the terminologies present in the enterprise glossary or dictionary and appropriate meanings are attached to them. The Apache Lucene search, a text search engine written in Java, is

---

[8] http://armandbrahaj.blog.al/2009/04/14/list-of-english-stop-words/

```
Heading
    XYZ

Keywords
    CH2M Consulting
    ABCJ Limited
    XYZ

Cardinal Text
    Null

Terms
    information processing solutions
    umbrella
    IT industry
    internet
    remote access
    document management
    document retrieval
    methods
    software
    enterprise information management
    area
    services
    consulting services and web
    world
    sizes
    organizations
    customers
    world-class benefits
    bonus programs
    competitive salaries
    company

Key Phrases
    include organizations of all sizes throughout the world, such as ABCJ Limited, CH2M Consulting, etc.

General Phrases
    covers many information processing solutions.
    is an IT industry and their umbrella
    provides remote access via the internet.
    offers several methods of document retrieval and document management and
    hosting services in the area of enterprise information management.
    offer consulting services and web
    offers competitive salaries, bonus programs and world-class benefits.
```

Figure 8-18: Result from TEXT2RDF Application

used in this research for matching the BOP with the domain specific terminologies of the enterprises. The details about Apache Lucene are provided in **Appendix E**.

The next two sections present the results obtained in annotation process and evaluation measures calculated for measuring the efficacy of proposed approach.

XYZ

It is an **information industry** and their umbrella covers many **information processing** solutions.

Their **software** offers several methods of **document retrieval** and **document management** and provides remote access via the **internet**.

XYZ customers include organizations of all sizes throughout the world, such as **ABCJ Limited**, **CH2M Consulting**, etc.

The company offers highly competitive salaries, **bonus payments** and world-class benefits.

List of terms (present in EUROVOC)

- Information Industry
- Information Processing
- Software
- Document retrieval
- Document Management
- Internet
- Bonus payment

List of keywords

- XYZ
- ABCJ Limited
- CH2M Consulting

Figure 8-19: Hyperlinked Document

8.5.1.   *Keyword Annotations*

In order to test the accuracy and precision of the proposed approach, the EUROVOC thesaurus is used in this research as a domain specific glossary. EUROVOC is a multilingual glossary maintained by Office for Official Publications of the European Communities (European Union publications office 2005). This glossary contains 6645 concepts from 21 domains with 3636 associative relationships and it is published in 23 languages. In this research, the SKOS version of EUROVOC is used as a standard glossary. SKOS (Simple Knowledge Organization Schema), as discussed in Section 8.2, is a semantic data model built upon semantic RDF and RDFS and used for sharing and linking knowledge organization systems such as thesauri, classification schemes, taxonomies, and any other type of structured controlled vocabularies (Miles, et al. 2005). The concepts available in this thesaurus are used for annotating the terms extracted from the documents.

The list of terms and phrases extracted from the sample text shown in Figure 8-17, using the TEXT2RDF application is provided in Figure 8-18. As evident from figure 8-18, the proposed approach classifies the BOP extracted the text document in different
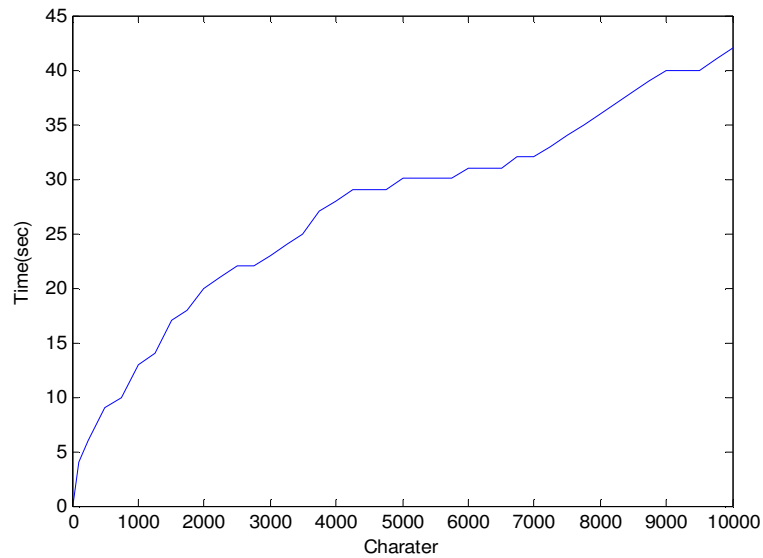
Figure 8-20: Computation time vs. Number of characters

categories, i.e. Heading, Keyword, Cardinal, Term, Key Phrase (IPhrase, in Chapter 6.4.3) and General phrases. The terms and phrases provided in the list are compared with the domain specific glossary of companies and local meanings are attached with the document. In this research, only terms and keywords are considered and compared with the EUROVOC thesaurus and ECOS profiles. The terms and keywords present in the thesaurus and ECOS profiles are annotated with their corresponding hyperlinks. The semantic document with hyperlinked keywords and terms is shown in Figure 8-19.

8.5.2. *Computational Time and Accuracy*

In order to test the applicability and robustness of the proposed TEXT2RDF approach, the application has been tested on two experimental datasets. The first dataset consists of enterprise profiles ranging from 500 to 2000 characters, whereas the other dataset consists of project reports varying from 2000 to 10000 characters. All experiments are performed in Java language and implemented on a HP personal computer with Pentium Central Processing (CPU) at 3.4 GHz.

In the literature, many different measures are proposed and used for measuring the performance of information retrieval systems (Hua Jiang 2009). However, computational time and accuracy, which are the two most crucial factors for any natural language processing technique, are used in this research. The computational time is the amount of time required to extract the bag of phrases from a text document and generate a RDF document from it. In this case study, 25 documents (20 enterprise profiles with <2000 characters and 5 projects reports with <10000 characters) were used for the analysis. The time consumed by the TEXT2RDF application in converting documents (ranging from 500-10000 characters) into a semantic resource is shown in Figure 8-20. From the figure, it is evident that computer usage is high; but the increase in computation time is polynomial which makes the approach useful and realistic.

The accuracy of the TEXT2RDF application was evaluated using precision (usefulness of BOP), recall (completeness of BOP) and F-measure (harmonic mean of precision and recall). Precision and recall are the most common measures used to evaluate the efficiency and comprehensiveness of information retrieval systems. In the research, these measures are calculated as:

$$Precision = \frac{|\{retrieved\ BOP\} \cap \{relevant\ BOP\}|}{|\{retrieved\ BOP\}|} \qquad 8\text{-}1$$

$$Recall = \frac{|\{retrieved\ BOP\} \cap \{relevant\ BOP\}|}{|\{relevant\ BOP\}|} \qquad 8\text{-}2$$

$$F = 2 \times \frac{precision \times recall}{recall + precision} \qquad 8\text{-}3$$

where, $retrieved\ BOP$ is the list of terms and phrases extracted from the text content using TEXT2RDF application, as shown in Figure 8-18; and $relevant\ BOP$ is the revised list of BOP obtained after analyzing it manually and comparing it with the results extracted manually from the text content. As mentioned earlier, the analysis of BOP and revision of RDF document is carried manually in this research

Table 8-9: Experiments results from two datasets

Type	Dataset 1: Enterprise Profile			Dataset 2: Project Reports		
	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure
Term	89.60	95.20	92.32	84.60	93.20	88.69
Keyword	92.40	100.00	96.05	86.10	98.40	91.84
Cardinal	100.00	100.00	100.00	96.40	98.80	97.59
Verb	81.10	100.00	89.56	75.60	94.50	84.00
Sentiment	72.20	88.30	79.44	66.60	80.70	69.87
Phrase	76.20	96.10	85.00	79.60	91.20	85.01
IPhrase	87.10	94.80	90.79	81.30	89.70	85.29
Average	85.51	96.34	90.45	81.46	92.36	86.57

For example, the analysis of BOP list shown in Figure 8-18 is as follows:

Term "consulting services and web" is not extracted correctly.

Term "web hosting services" is not in the BOP. As hosting is tagged as Verb by NLP parser, this term is not considered as a single phrase.

Term "umbrella", "world", "sizes" and "area" are not relevant to an IT industry.

Similarly, the TEXT2RDF was used to extract BOP from the two experimental document sets (enterprise profiles and project reports) and the results were analysed manually. The average results obtained from 25 documents (20 profiles and 5 reports) are provided in Table 8-9. From table 8-9, it is clear that the recall values are promising for both kinds of document sets, which confirm that no information is neglected from the text document. However, the precision values vary for different types of labels in BOP.

In this research, the focus was to develop an information retrieval system for extracting domain specific meanings from the enterprise documents, and the scope was limited to the corporate memory of a single enterprise. Therefore, in order not to lose any type of term or phrase from the enterprise document, a stop word list with a limited set of words was used in this application. Apparently, the results with higher recall show that TEXT2RDF application extracts all the terms and phrases present in the text content. However, it also extracts irrelevant terms and phrases from the enterprise documents,

which then lowers the precision measure. The other obvious reasons observed for lower precision and recall in extracting BOP are as follows:

• *Term and Keyword*: The conjunction 'and' used, though seldom, within the word, e.g. McCone and Company, health and safety etc.

• *Verb*: The terms having two or more words such as "web hosting services", with verb in it are not extracted correctly. Also, the sentence having more than one verb drops off the precision of the proposed approach.

• *Phrase and Iphrase*: The stop word filtration process utilized in this approach sometimes filters relevant information from the phrases, thus reducing its correctness.

8.6. Termination Stage

This is the stage when the collaborative project is finished and the virtual enterprise is disbanded. At this stage, each enterprise archives their project related information and updates their knowledge base with the lessons learnt from a project. As mentioned in Chapter 7, companies generally maintain a corporate memory to explicitly store the collective knowledge possessed by human capital and technology, organizational structures and organizational culture and makes it available to the entire company. A semantic framework is proposed in this research for corporate memory management on the semantic web. The implementation of the proposed framework on a real life case-study is shown in this section.

8.6.1.   *Experimental Setup*

As discussed in Chapter 7.2, corporate memory takes into account the diversity of knowledge and information found in an organization. Dieng Kuntz et al. distinguished internal memory from the external memory in an organization (Dieng-Kuntz, et al. 2001). Tourtier proposed four types of corporate memory, i.e. profession, society, individual and project (Tourtier 1995). In this research, two types of corporate memory are considered as a case study, i.e. profession memory and project memory, explained as follows:

- *Professional Memory*: This captures skills, capability, expertise and competence of that an organization is able to offer. It defines the skills and abilities based on knowledge and experience of its human capital, methods and resources. This information is utilized to coordinate the deployment of a company's assets in ways that give competitive advantage and produce success in the market place. In the past, professional memory was handled by humans having good knowledge about competences and requirements of the company. In the current knowledge economy, it is really difficult to handle such huge amounts of information and thus it is necessary to model knowledge in a form that is manageable and addressable. In this

research, the proposed approach is used for storing profession memory as a semantic web resource. To demonstrate this implementation, an experimental dataset has been taken from Yahoo Business (http://biz.yahoo.com). The experimental data set consists of brief descriptions of 20 companies from 4 different sectors. The length of summaries varies from 500 to 2000 characters, averaging 1150 characters per description.

• *Project Memory*: This type of corporate memory stores information about projects handled or completed by an organization. Project documents capture the project experience- both good and bad. It is a rich source of knowledge and data for organizations- if organizations have the time and resources to analyze them. In general, these reports are analyzed to get important details and learn lessons from past experience. The results are then used to enhance processes, improve customer relationships, identify specific problem areas etc. However, most organizations lack resources to examine this memory and thus miss important insights thereby leading to a missed opportunity to learn from past projects. The proposed concept of semantic corporate memory is utilized here to extract knowledge from text based reports. To demonstrate this approach, 5 project reports were considered as an experimental data set. These reports vary from 2000 to 10000 characters, averaging 4500 characters per description.

In the next section, these two experimental datasets are analyzed and relations are identified among them (separately) using the corporate memory management framework proposed in Chapter 7.

8.6.2.   *Semantic Corporate Memory Management*

The corporate memory management framework, explained in Section 7.3, analyzes a set of enterprise documents by extracting information from their text contents and then identifying relations among the documents. The proposed framework carries out this analysis in four steps, i.e. 1) Converting enterprise document in plain text format, 2)

Mining information from the text content, 3) Identifying relations among the text content, and 4) Storing these relations for future use.

In order to realize this framework, a Java based program is coded in this research. At first, the TEXT2RDF web application, created in previous section 8.5, is used to extract a list of terms and phrases (BOP) from the text documents. Subsequently, the BOP from a set of documents is pooled together and relations are identified among them. This analysis is carried out separately and differently for the terms and phrases present in BOP. However for brevity and clear illustration of the application, only the terms and keywords present in the BOP are considered in this case study and relations are identified among them.

The latent semantic analysis (LSA) technique, discussed in Section 7.3.3.1, is used in this research for identifying relations among the terms and keywords present in BOP. In this analysis, a term-document matrix is created with each cell representing the frequency of the term in a document. Then, the matrix is decomposed using the SVD (singular value decomposition) technique and the result matrices are used for generating the Document-Term graph. The step-wise implementation of the LSA technique on a small-size problem is explained in Section 7.3.3.1. In this section, only the document-term graphs generated for the two experimental data sets are provided.

A software JAMA[9], Java Matrix Package, is used in this programme for decomposing document-term matrix using SVD, details are in **Appendix E**. The graphs are plotted using MATLAB 6.1[10].

The document-term graphs generated from the two experimental datasets are follows:

---

[9] http://math.nist.gov/javanumerics/jama/

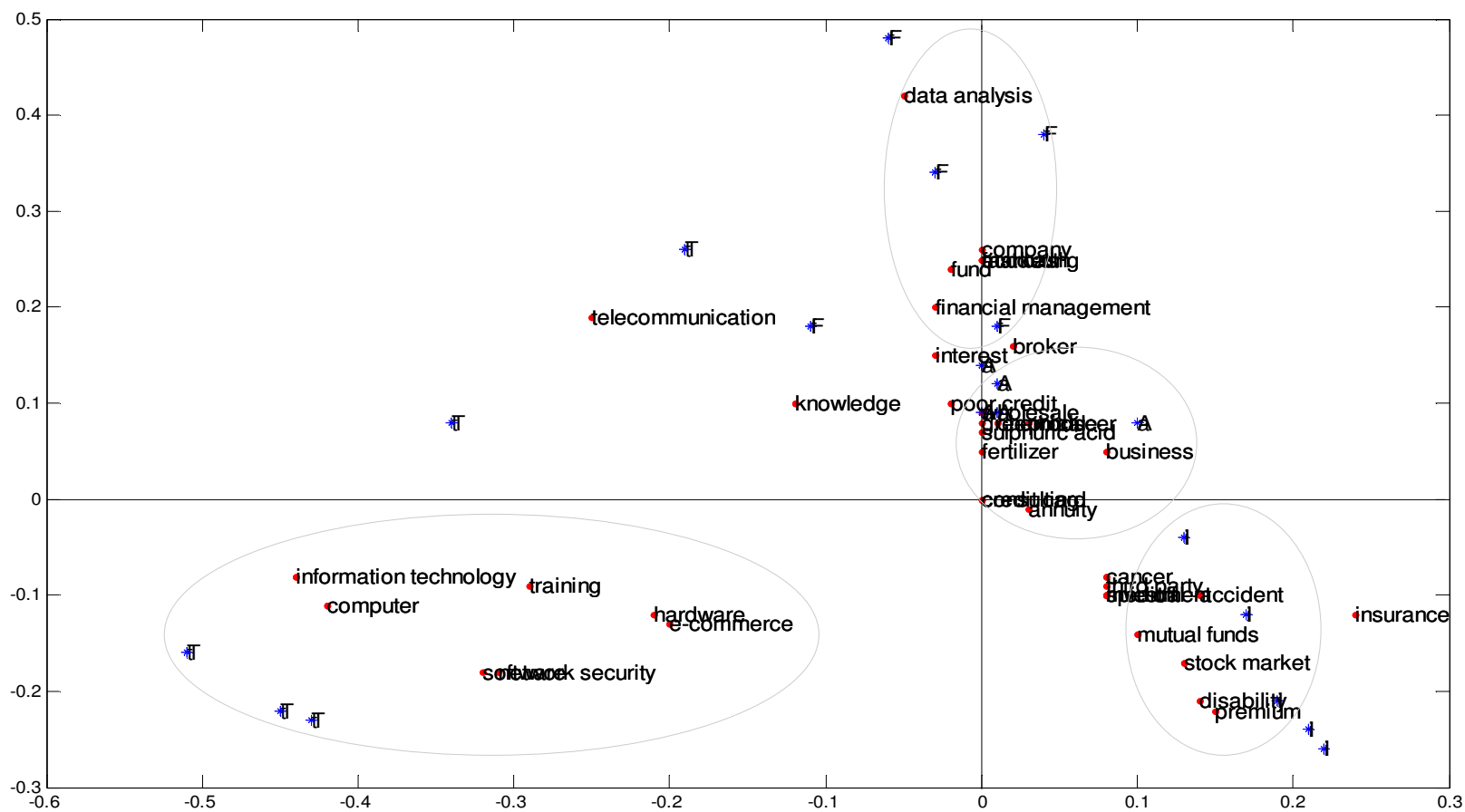[10] http://www.mathworks.com/products/matlab/

Figure 8-21: Document-term graph for Enterprise Profiles

- *Yahoo Profile*: At first, the terms and keywords were extracted from the documents present in the first dataset having 20 enterprise profile. The terms and keywords extracted from the documents were pool together and the abovementioned LSA technique was implemented on them. The document-term graph between 20 documents and 40 terms selected from BOP is shown in Figure 8-21. The documents in graph are labelled with single letters (i-insurance, t-technology, f-finance, a-agriculture) and words are labelled with their name. After analyzing the graph and measuring the distance of each term from each document, it was observed that terms and keywords are plotted close to their domain specific documents. For example, the term "information technology" is close to the documents from technology domain (labelled as T), whereas the term "stock-market" is closer to the documents from insurance sector. The four clusters are marked in the graph to clearly show that sector specific keywords are placed close to the documents.

- *Project Report*: Similarly, the project reports were analysed by extracting the terms and keywords from the text content and generating a document-term graph from them. However, as the number of terms and keywords extracted from the document were quite high, the distance of each term from the documents were measured and analysed. On analysing the measured distance, it was observed that each document includes mainly three types of terms and keywords, i.e. template specific (T)– present in all documents, company specific (C)- often used in project documents, project specific (P)- used in particular document. For example, the terms like "health and safety" and "contract period" were in all the reports, whereas the keywords like "Phil Taylor" and "South Wales" were in one or few reports. The document-terms graphs generated with these three types of terms and keywords is
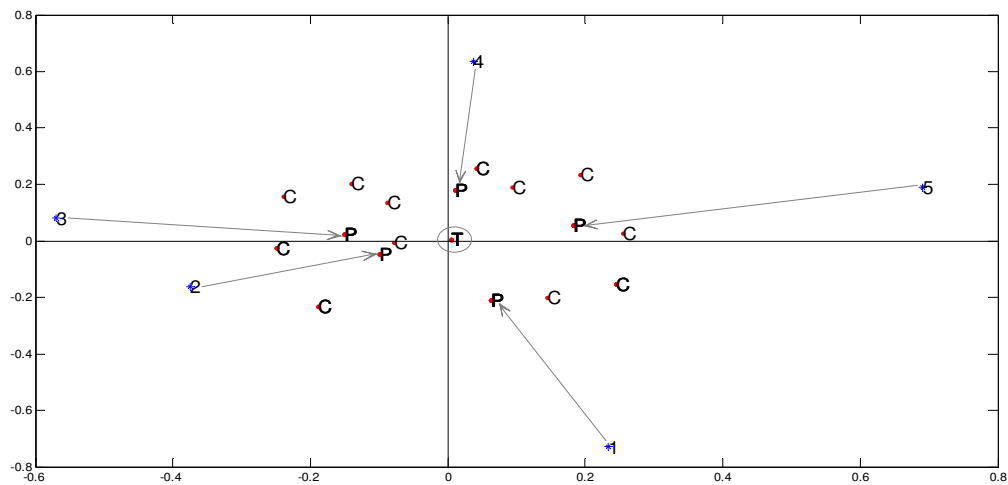
Figure 8-22: Document Term graph for Project Reports

(C- Company Specific, T- Template Specific and P- Project Specific) (1-5 are Project Report Numbers)

shown in Figure 8-22, where documents are labelled as 1-5 and the three types of terms and keywords are labelled as 'C', 'P' and 'T'. These clusters of keywords obtained from the reports can be stored separately and utilized in enterprise knowledge management.

8.7. Concluding Remarks

Information sharing and knowledge management are major requirements for the effective and efficient functioning of the virtual enterprise. In a collaborative environment, it is necessary to maintain interoperability and preserve syntactic and semantic content during the exchange of messages. In this research, emerging semantic technologies, i.e. ontology modelling, semantic web services and information retrieval techniques are utilized for facilitating the meaningful acquisition, sharing and reuse of information resources. A semantic framework is proposed for each stage of virtual enterprise to provide explicit semantic description to the information resources and facilitate their sharing and management among heterogonous enterprise applications.

In this chapter, an exemplar case study is considered to demonstrate the applicability and effectiveness of proposed semantic frameworks. A prototype implementation of the proposed frameworks is developed for each stage of virtual enterprise and a detailed description on an example virtual enterprise is provided in this chapter. In order to illustrate the proposed concepts lucidly and to extend the work in a real life scenario, a web application is developed in this research. The snapshots from the web application are presented in this chapter and entire web application is available at http://kmm.lboro.ac.uk/projects/phd#NiteshKhilwani.

C h a p t e r   9

CONCLUSIONS

The road to success is always under construction.

Lily Tomlin

ABSTRACT

The investigations performed in this research thesis have led to many interesting observations and new insights, which are summarized in this chapter. The key contributions made to achieve the aims and objectives of this research are also provided in this chapter. Finally, the conclusion is provided by presenting some interesting directions for future research.

CHAPTER OUTLINE

9.1.   Research Objectives and Contributions

9.2.   Deepening of the research

9.3.   Future Scope

9.1. Research Objectives and Contributions

Information and knowledge are the most valuable organizational assets in a virtual collaborative environment. These intangible resources provide power in many of the enterprise functions enabling and facilitating the preservation of valuable heritage, learning new things, solving intricate problems, and creating core competencies. However, in a virtual collaborative environment information resources are generally distributed within partners, rather than being centralized. Accessing heterogeneous and distributed information resources in a coordinated and collaborative way requires adequate syntactic and semantic understanding to get a seamless access and retrieval of the right information resources, and reduce the risk of misinterpretation.

Considering the above requirements, the main objective of this research was defined as: *To analyse the life cycle of a virtual enterprise, identify problems and requirements for information sharing and knowledge management tasks in each stage of virtual enterprise and propose semantic web framework and related web services to facilitate those tasks.* This objective was achieved by satisfaction of the 4 aims discussed in Chapter 3. The research contributions made for each objective is as follows:

1. *Analyze the life cycle of virtual enterprise and identify information sharing and knowledge management issues in each stage.*

- The concept of *Virtual Enterprise* is one of the most competitive approaches for independent enterprises, which enables them to cooperate and share heterogeneous information resources without losing their autonomy and independence. After analyzing six different models of virtual enterprise, discussed in Table 2-2, the life cycle of a virtual enterprise was divided into four stages, namely: identification, creation, operation and termination. The key goals and processes involved in each stage of virtual enterprise for six different collaboration models were provided in Table 2-3. The key issues associated with information sharing and knowledge

management were identified for each stage of a virtual enterprise and a semantic framework with tools and techniques were proposed to resolve them.

- ***Conclusion 1***: After analysing the life cycle of virtual enterprise and identifying key goals and processes in stage for different styles of collaborative networks, it is concluded that information sharing and knowledge management is a key issue in all the life stages of an open networked virtual collaboration.

2. *Explore the need of explicit semantic description and machine understanding in information sharing and knowledge management at each stage of enterprise collaboration.*

- A basic requirement for a collaborative network is that companies must be able to interoperate and exchange information in real time so that they can work as a single integrated unit whilst keeping their independence and autonomy. As part of an effort to improve cooperation, communication and collaboration, sets of interconnected data and semantic models are proposed in this research. This research has mainly focused on developing semantic agreements in each stage of the virtual enterprise life cycle by fixing standardized meanings and relations of terminologies used by them.

- ***Conclusion 2***: In order to reduce the risk of misinterpretation in understanding heterogeneous and distributed information resources in a collaborative networked environment, explicit semantic descriptions are required for enterprise information resources. After studying recent research work and market requirements, it is concluded that ontology models and semantic web technology should be utilized to enable computers to automatically process and understand the information which should be shared within the collaborative network.

3. *Develop framework to facilitate enterprise collaboration with emerging technologies, e.g. semantic web, ontology and information retrieval techniques.*

- This research work has exploited the emerging semantic technologies e.g. ontology modelling, semantic web and text mining to facilitate information sharing and knowledge management in enterprise collaboration. The purpose of introducing semantic web technology is to enhance the effectiveness of communication and collaboration infrastructure necessary for the capture and reuse of information. The semantic frameworks proposed for each stage of virtual enterprise are:

  i). *Identification*: ECOS ontology model was developed in Chapter 4 for capturing the published competences of enterprises.

  ii). *Creation*: The semantic profiles based on ECOS ontology were analyzed in Chapter 5, for mapping and measuring the relations between enterprises and selecting suitable business opportunities and partners.

- ***Conclusion 3***: The design and implementation of ECOS ontology model and its utilization and testing for identifying relations among enterprise profiles, clearly demonstrates that an ontology model for enterprise competences can facilitate the process of publishing, sharing, managing and relating enterprise profiles over the web.

  iii). *Operation*: An information retrieval tool, TEXT2RDF, was proposed in Chapter 6 for extracting information from the text documents and annotating with domain specific meanings.

  iv). *Termination*: A semantic corporate memory framework was proposed in Chapter 7 for identifying relations between documents present in enterprise information repository.

- ***Conclusion 4***: The TEXT2RDF application designed and implemented for information extraction and semantic annotation can be used to represent text documents in formalisms that both machines and humans can understand and perform intelligent search, querying and reasoning on them.

4. *Evaluate the quality and feasibility of proposed framework on an experimental or real life case study.*

- The quality and feasibility of semantic frameworks proposed in Chapter 4 to 7 was validated and evaluated in Chapter 8 of this thesis. An experimental case study with 25 enterprises was considered to demonstrate the real life implementation of the tools and techniques proposed in this research.

- **Conclusion 5**: From the implementation of the proposed semantic frameworks on a real life case study, it is concluded that companies can utilize them in facilitating various enterprise applications related with information sharing and knowledge management, and also include them in their existing software applications.

Based on the contributions to the research questions, the achieved objectives and concluding remarks, this research contributes to theoretical, methodological and practical body of knowledge.

9.2. Deepening of the research

A Virtual Enterprise is often visualized as a next generation enterprise, which consists of a network of globally distributed firms operating in an environment of abrupt and often unanticipated change. It is an industrial environment comprised of any number of disparate companies and heterogeneous applications having their own local information and knowledge resources, which are often distributed in different geographical locations. The current market need is to develop infrastructures that effectively and efficiently manage the distributed information resources and enables enterprises to interact and collaborate intensively. A general requirement for the infrastructure is to support virtual collaborations, and it is therefore required that enterprises must be able to inter-operate and exchange information in real time so that they can work as a single integrated unit, while keeping their independence and autonomy.

The developments in communications and computer networking infrastructures are enabling factors for distributed companies to collaborate electronically and form virtual collaborations. However, the legacy systems running at present in the enterprises have not been developed to directly and independently connect with the systems in other enterprises. Moreover, the enterprise information resources are not structured with standard mechanisms, which can be readable and understandable to both humans and machines without any risk of misinterpretation. In order to support the information sharing and knowledge management with less or without human involvement, this research presents semantic frameworks to describe the resources in a way that is understandable and usable by the networked organization. The two types of information resources considered in this research are:

1. *Competence Profile*: The competence profiles are generally published by the companies willing to communicate and collaborate with other companies, and exploit their skills and capabilities in the global market. ECOS ontology was developed in this research as a standard metadata for publishing the competences in a machine readable format. The ECOS profiles are further utilized in automating various enterprise applications, i.e. network analysis, business opportunity selection, partner selection and virtual enterprise formation.

2. *Text Documents*: Enterprises often share documents in a virtual collaboration to satisfy the request of the partners and requirements of the projects. These documents often contain domain specific terminologies, which hamper the sharing and understanding of the contents of the documents. A TEXT2RDF application is proposed in this research to extract the information from the documents and annotate with their explicit definitions. Later, the semantic content of the documents are analyzed to identify relations between terminologies present in same document and with other documents.

In this research, semantic frameworks are developed to convert the abovementioned two types of documents into a semantic web resource. Similarly, other information

resources can be considered and advanced concepts can be proposed in the future research.

9.3. Future Scope

The section outlines directions of future work related to the results presented in this thesis that were not in the scope of this work:

- *Virtual Enterprise Networks*: This research work started with the study of six different types of virtual collaborations, but later developed a generic life cycle for virtual enterprise and restricted with its analysis. In future work, different types of virtual networks should be studied and their problems can be handled.

- *Information sharing and Knowledge Management*: The focus of this research was mainly on analysing the life cycle of a virtual enterprise and solving the problems pertaining to the information sharing and knowledge management. However, there are various other problems in virtual collaborations, such as information handling, information storage, knowledge acquisition, knowledge dissemination etc.

- *Ontology Development*: Ontologies are used for defining the semantics present in the concerned domain and for sharing them among people, databases and applications. In this research, the ECOS ontology was developed as a standard metadata for capturing enterprise competences and publishing them in a computer understandable format. In order to avoid reinvention of the wheel, existing ontologies (e.g. vCard) and standard terminologies (e.g. CPC, ISIC, etc) were used for explaining certain ECOS concepts. In future, this ontology can be reused in bigger enterprise ontologies for capturing the top level information of enterprises.

- *Ontology Population*: Ontologies are nice to look at, but it is very difficult to integrate and populate them with the right information resources. In general, this task is handled by domain experts, but it is a lengthy and costly process. For example, the current version of ECOS ontology requires manual effort to fill the information in a web form and create ECOS-card for the enterprises. To reduce the effort of

populating an ontology, a specialized knowledge service can be developed that can search and extract specific information (such as address, contact person, etc) directly from the text documents explaining industrial or individual profiles. Further, the automated process can be used to create an ECOS-card automatically from existing web pages or any other information resource from companies.

- *Enterprise Network Analysis*: It is a very active field of research in Web 2.0, the social web, in which the focus is on studying communication and socio-technical networks among organizations. In this research, enterprise networks and relationships were identified using the ECOS ontology model and were later used in suggestion virtual collaboration. This analysis can be further extended with other attributes, complex mathematical formulations and advanced algorithms.

- *Information Retrieval and Semantic Annotation*: These are considered as the key technologies in the development of Web 3.0, the next generation web. In Web 3.0, the idea is to add meaning to the information on the web and generate *an intelligent web* by using machine learning, intelligent applications, natural language processing, etc. In this research, the TEXT2RDF application was developed to extract information from the documents and annotate with domain specific meanings for reducing the risk of misinterpretation in enterprise collaboration. In future, this tool can be used in automatic extraction and semantic annotation of information resources, and can also be modified for higher rate of accuracy and lower computation time.

- *Corporate Knowledge Management*: In this research, the idea behind developing semantic corporate knowledge was to link well defined meanings with the content of the information resources and utilize them in identifying relations within and among them. However, the semantic framework developed was limited only for the text documents and in future can be advanced for other types of enterprise information resources.

REFERENCES

AHUJA, M.K. and CARLEY, K.M., 1999. Network Structure in Virtual Organizations. Organization Science, 10(6), 741-757.

ALANI, H., KIM, S., MILLARD, D.E., WEAL, M.J., HALL, W., LEWIS, P.H. and SHADBOLT, N.R., 2003. Automatic ontology-based knowledge extraction from web documents. IEEE Intelligent Systems, 18(1), 14-21.

ALLEN, C., 2003. Competencies. 1. HR-XML Consortium, Available:http://ns.hr-xml.org/2_0/HR-XML-2_0/CPO/Competencies.pdf.

ANKLAM, P., 2003. Tapping social networks to leverage knowledge and innovation, 7 May 2003, KnowledgeNets, New York,

ARKIN, A., 2002. Business Proces Modelling Language, Available: http://xml.coverpages.org/BPML-2002.pdf.

BARRAT, A., BARTHALEMY, M., PASTOR-SATORRAS, R. and VESPIGNANI, A., 2004. The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America, 101(11), 3747-3752.

BERNERS-LEE, T., JAMES, H. and ORA, L., 2001. The semantic web. Scientific American Magazine, Available: http://www.scientificamerican.com/article.cfm?id=the-semantic-web

BERRE, A., HAHN, A., AKEHURST, D.H., BEZIVIN, J., TSALGATIDOU, A., VERMAUT, F., KUTVONEN, L. and LININGTON, P.F., November 2004. State-of-the art for Interoperability architecture approaches. D9.1.

BIZER, C., HEATH, T. and BERNERS-LEE, T., 2009. Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, 5(3), 1-22.

BLOCH, M., 1999, Web 2.0, mashups and social networking - what is it all about?. Available: http://www.tamingthebeast.net/articles6/web2-mashups-social-network.htm.

BOUCHER, X., BONJOUR, E. and GRABOT, B., 2007. Formalisation and use of competencies for industrial performance optimisation: A survey. Computers in Industry, 58(2), 98-117.

BRADFORD, R., 2008. Why LSI? Latent Semantic Indexing and Information Retrieval, Content Analyst. White Paper edn. Content Analyst Company, LLC, Available: http://www.contentanalyst.com/images/images/overview_lsi_v1.pdf.

BRANDES, U., 2001. A Faster Algorithm for Betweenness Centrality. Journal of Mathematical Sociology, 25, 163-177.

BRANDT, S.C., MORBACH, J., MIATIDIS, M., THEIßEN, M., JARKE, M. and MARQUARDT, W., 2008. An ontology-based approach to knowledge management in design processes. Computers & Chemical Engineering, 32(1-2), 320-342.

BRICKLEY, D. and MILLER, L., 2007. FOAF Vocabulary Specification 0.9. Available: http://xmlns.com/foaf/spec/

BROWNE, J. and ZHANG, J., 1999. Extended and virtual enterprise- similarities and differences. International Journal of Agile Management Systems, 1(1), 30-36.

BURN, J., MARSHALL, P. and WILD, M., 1999. Managing knowledge for strategic advantage in the virtual organisation, 1999, ACM pp19-26.

CAMARINHA-MATOS, L.M. and AFSARMANESH, H., 2005. Collaborative networks: a new scientific discipline. Journal of Intelligent Manufacturing, 16(4), 439-452.

CAMARINHA-MATOS, L.M., AFSARMANESH, H., GARITA, C. and LIMA, C., 1998. Towards an architecture for virtual enterprises. Journal of Intelligent Manufacturing, 9(2), 189-199.

CANAVESIO, M.M. and MARTINEZ, E., 2007. Enterprise modelling of a project-oriented fractal company for SMEs networking, Computers in Industry, 58(8-9), 794-813.

CAO, H.Y. and WANG, D.W., 2002. Genetic algorithm for a multi-objective optimization model of partner selection in virtual enterprise. Control Decision, 17(1), 274-277.

CASEY, M. and PAHL, C., 2003. Web Components and the Semantic Web. Electronic Notes in Theoretical Computer Science, 82(5), 156-163.

CASTANO, S., FERRARA, A. and MONTANELLI, S., 2006. Ontology-based Interoperability Services for Semantic Collaboration in Open Networked Systems. In: D. KONSTANTAS, J. BOURRIÈRES, M. LÉONARD and N. BOUDJLIDA, eds, Interoperability of Enterprise Software and Applications. London, UK: Springer, pp. 135-146.

CERAMI, E., 2002. Web services essentials. O'Reilly Media.

CHIESA, V. and MANZINI, R., 1997. Competence levels within firms: a static and dynamic analysis. In: A. HEENE and R. SANCHEZ, eds, Competence-based strategic management. Wiley: Chichester, pp. 195-214.

CHURCH, K.W. and MERCER, R.L., 1993. Introduction to the Special Issue on Computational Linguistics Using Large Corpora. Computational Linguistics, Special Issue on Using Large Corpora I, 19(1), 1-24.

CORCHO, O., FERNANDEZ-LOPEZ, M. and GOMEZ-PEREZ, A., 2003. Methodologies, tools and languages for building ontologies. Where is their meeting point?. Data & Knowledge Engineering, 46(1), 41-64.

CROSS, R. and PRUSAK, L., 2002. The people who make organizations go--or stop. Harvard Business Review, 80(6), 104-112.

CROSS, R., THOMAS, R., DUTRA, A. and NEWBERRY, C., 2007. Using Network Analysis To Build a New Business. Organizational Dynamics, 36, 345-362.

DANG, H.T., KIPPER, K., PALMER, M. and ROSENZWEIG, J., 1998. Investigating regular sense extensions based on intersective Levin classes, 1998, Association for Computational Linguistics pp293-299.

DAVALCU, H., VADREVU, S., NAGARAJAN, S. and RAMAKRISHNAN, I.V., 2003. OntoMiner: bootstrapping and populating ontologies from domain-specific Web sites. Intelligent Systems, IEEE, 18(5), 24-33.

DAVIDOW, W.H. and MALONE, M.S., 1992. The Virtual Corporation – Structuring and Revitalizing the Corporation for the 21st Century. New York: HarperCollins.

DENG, H., 1999. Multicriteria analysis with fuzzy pairwise comparison, Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE '99. 1999 IEEE International, August 22-25 1999, pp726-731.

DIENG, R., CORBY, O., GIBOIN, A. and RIBIÈRE, M., 1999. Methods and tools for corporate knowledge management. International Journal of Human-Computer Studies, 51(3), 567-598.

DIENG-KUNTZ, R., CORBY, O., GANDON, F., GIBOIN, A., GOLEBIOWSKA, J., MATTA, N. and RIBIÈRE, M., 2001. Methods and tools for knowledge management: A multidisciplinary approach to Knowledge Management. 2nd edn. France: Microsoft Press.

DRAGANIDIS, F., CHAMOPOULOU, P. and MENTZAS, G., 2006. An Ontology Based Tool for Competency Management and Learning Paths, 6th International Conference on Knowledge Management, I-KNOW, Graz, Austria.

ECKERSON, W.W., 1995. Three tier client/server architecture: Achieving scalability, performance, and efficiency in client server applications. Open Information Systems, 10(3), 1-20.

ECOLEAD, June 25, 2008, 2008-last update, European Collaborative Networked Organisations Leadership Initiative. Available: http://ecolead.vtt.fi/ [7/14, 2008].

EL-DIRABY, T.E. and ZHANG, J., 2006. A semantic framework to support corporate memory management in building construction. Automation in Construction, 15(4), 504-521.

ELLIMAN, D. and PULIDO.JRG., 2002. Visualizing ontology components through self-organizing maps, 2002, IEEE Computer Soc. Press pp434-438.

ERDMANN, M., MAEDCHE, A., SCHNURR, H. and STAAB, S., 2000. From manual to semi-automatic semantic annotation: About ontology-based text annotation tools, July 31 - August 4 2000, Morgan Kaufmann.

EUROPEAN UNION PUBLICATIONS OFFICE, 2005. EUROVOC thesaurus, Available: http://eurovoc.europa.eu/.

FADEL, F., FOX, M.S. and GRUNINGER, M., 1994. A generic enterprise resource ontology. In Proceedings of the Third Workshop on Enabling Technologies - Infrastructures for Collaborative Enterprises, 86-92.

FENG, W., CHEN, J. and ZHAO, C., 2000. Partners selection process and optimization model for virtual corporations based on genetic algorithms. Journal of Tsinghua Univ. Science Technology, 40(10), 120-124.

FORBAIRT, 1996. Virtual Corporation Defined, Summary Section for Forbairt Internet Report. Ireland: Forbairt.

FOX, M.S., 1993. Issues in Enterprise Modelling, 1993, IEEE pp 86-92.

FRAGIDIS, G., PASCHALOUDIS, D. and TSOURELA, M., 2008. Towards an Educational Model for the Knowledge Economy. Communications of the IBIMA, 3(9), 62-67.

FREEMAN, L.C., 1979. Centrality in social networks conceptual clarification. Social Networks, 1(3), 215-239.

FUQING, Z., YI, H. and DONGMEI, Y., 2006. A multi-objective optimization model of the partner selection problem in a virtual enterprise and its solution with genetic algorithms. The International Journal of Advanced Manufacturing Technology, 28(11), 1246-1253.

GHODSYPOUR, S.H. and O'BRIEN, C., 2001. The total cost of logistics in supplier selection, under conditions of multiple sourcing, multiple criteria and capacity constraint. International Journal of Production Economics, 73(1), 15-27.

GIALELIS, J.V., KALOGERAS, A.P., ALEXAKOS, C.E. and PAPADOPOULOS, G., 2005. Manufacturing Collaborative Process Integration Utilizing State of the Art Technologies. Industrial Electronics, 2005. ISIE 2005. Proceedings of the IEEE International Symposium on, 4, 1429-1434.

GOLBECK, J. and ROTHSTEIN, M., 2008. Linking social networks on the web with FOAF: a semantic web case study, 13-17 July 2008, AAAI Press pp1138-1143.

GÓMEZ-PÉREZ, A. and SUÁREZ-FIGUEROA, M.C.E.-., 2004. Evaluation of RDF(S) and DAML+OIL Import/Export Services within Ontology Platforms.

GOOSSENAERTS, J.B.M. and PELLETIER, C., 2002. Ontology and Enterprise Modelling. In: Eijnatten. Participative Simulation Environment for Integral Manufacturing Enterprise Renewal. TNO Arbeid, 41-52.

GOTTGTROY, P., KASABOV, N. and MACDONELL, S., 2006. Chapter 21 Evolving ontologies for intelligent decision support. In: ELIE SANCHEZ, ed, Capturing Intelligence, Elsevier, pp. 415-439.

GRANSIER, T. and SCHÖNEWOLF, W., 1995. Validation of CIMOSA. Computers in Industry, 27(2), 95-100.

GREY, D., 2003. Corporate memory - the hard way, Knowledge at work, Available: http://denham.typepad.com/km/2003/09/corporate_memor.html.

GRUBER, T.R., 1993. A translation approach to portable ontologies. Knowledge Acquisition, 5(2), 199-220.

GUARINO, N. and GIARETTA, P., 1995. Ontologies and knowledge bases: towards a terminological clarification. In Mars, N., editor, Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing, 25-32.

GUARINO, N., 1998. Formal Ontology in information systems. 1st International Conference on Formal Ontology in Information Systems, 3-15, Publisher: IOS Press, Place of Publication: Amsterdam, The Netherlands

GUHA, R. and MCCOOL, R., 2003. Tap: the semantic platform. Computer Networks, 42(5), 557-577.

HAJIDIMITRIOU, Y.A. and GEORGIOU, A.C., 2002. A goal programming model for partner selection decisions in international joint ventures. European Journal of Operational Research, 138(3), 649-662.

HALPIN, H., SUDA, B. and WALSH, N., 2006. An ontology for vCards, Available: http://www.w3.org/2006/vcard/ns.

HANDSCHUH, S. and STAAB, S., 2002. Authoring and annotation of web pages in CREAM, 2002, ACM pp462-473.

HANDSCHUH, S., STAAB, S. and CIRAVEGNA, F., 2002. S-CREAM — Semi-automatic CREAtion of Metadata. In: A. GÓMEZ-PÉREZ and V. BENJAMINS, eds, Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. Lecture Notes in Computer Science edn. Berlin/Heidelberg: Springer, pp. 165-184.

HARARY, F., 1969. Graph theory. Reading, MA: Addison-Wesley.

HARDING, J.A. and POPPLEWELL, K., 1999. Information modelling: an integration of views of a manufacturing enterprise. International Journal of Production Research, 37(12), 2777.

HARZALLAH, M., BERIO, G. and VERNADAT, F., 2006. Analysis and modelling of individual competencies: toward better management of human resources. Systems, Man and Cybernetics, Part A, IEEE Transactions on, 36(1), 187-207.

HEFLIN, J. and HENDLER, J., 2000. Searching the Web with SHOE, 2000, AAAI Press pp35-40.

HEXIN LV and BIN ZHU, 2006. Skill ontology-based semantic model and its matching algorithm.

HOUGHTON, J. and SHEEHAN, P., 2000. A Primer on the Knowledge Economy. Centre for Strategic Economic Studies, Victoria, Australia.

HSIEH, T., LU, S. and TZENG, G., 2004. Fuzzy MCDM approach for planning and design tenders selection in public office buildings. International Journal of Project Management, 22(7), 573-584.

HUA JIANG, 2009. Study on the Performance Measure of Information Retrieval Models, Intelligent Ubiquitous Computing and Education, 2009 International Symposium on, 2009, pp436-439.

HUANG, C., TSENG, T. and KUSIAK, A., 2005. XML-Based Modelling of Corporate Memory. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 35(5), 629-640.

HUANG, N. and DIAO, S., 2007. Ontology based enterprise knowledge integration. Robotics and computer integrated manufacturing, 24(4), pp 562-571.

HUTCHISON, T., 2008. Social Networking Sites. Web Marketing for the Music Business. Boston: Focal Press, pp. 201-225.

IANNELLA, R., 2001. Representing vCard objects in RDF/XML. W3C, Available: http://www.w3.org/TR/vcard-rdf.

IMS, 1996. Intelligent Manufacturing Systems, Available: http://www.img.org.

JONKERS, H., LANKHORST, M.M., BUUREN, R.V., HOPPENBROUWERS, S., BONSANGUE, M. and TORRE, L.V.D., 2004. Concepts for modelling enterprise architectures. International Journal of Cooperative Information Systems, 13(3), 257-287.

JUSSUPOVA-MARIETHOZ, Y. and PROBST, A., 2007. Business concepts ontology for an enterprise performance and competences monitoring. Computers in Industry, 58(2), 118-129.

KAHAN, J. and KOIVUNEN, M., 2001. Annotea: An Open RDF Infrastructure for Shared Web Annotations, May 1-5 2001, ACM pp623-632.

Kanet, J. J., Faisst, W., Martin, P., Application of information technology to a virtual enterprise broker: the case of Bill Epstein, International Journal of Production Economics 1999, Vol 62, pp 23-32.

KATZY, B., ZHANG, C. and L¨OH, H., 2005. Reference Models for Virtual Organisations. In: L.M. CAMARINHA-MATOS, H. AFSARMANESH and M. OLLUS, eds, Virtual Organizations. US: Springer, pp. 45-58.

KHAN, L. and LUO, F., 2002. Ontology construction for information selection, 14th IEEE International Conference on Tools with Artificial Intelligence, Crystal City, Virginia, pp122-127.

KHILWANI, N., HARDING, J.A. and CHOUDHURY, A.K., 2009. Semantic web in manufacturing. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 223(7), 905-924.

KIM, W., JEONG, O. and LEE, S., 2010. On social Web sites. Information Systems, 35(2), 215-236.

KLEIN, D. and MANNING, C.D., 2003. Accurate unlexicalized parsing, 2003, Association for Computational Linguistic pp423-430.

KOIVUNEN, M.R. and SWICK, R., 2001. Metadata Based Annotation Infrastructure offers flexibility and extensibility for collaborative applications and Beyond, 2001, ACM.

KOSANKE, K., 1995. CIMOSA-Overview and status. Computers in Industry, 27(2), 101-109.

KRACKHARDT, D. and HANSON, J.R., 1993. Informal networks: the company behind the chart. Harvard Business Review, 71(4), 104-111.

KUHN, O. and ABECKER, A., 1997. Corporate memories for knowledge management in industrial practice: prospects and challenges. Journal of Universal Computer Science, 3, 929-954.

KUTTNER, B., 1993. The pitfalls of virtual corporation with no permanent workers. The Boston Globe, syndicated Column.

LAI, L.F., 2007. A knowledge engineering approach to knowledge management. Information Sciences, 177(19), 4072-4094.

LAMMARI, N. and MÉTAIS, E., 2004. Building and maintaining ontologies: a set of algorithms. Data & Knowledge Engineering, 48(2), 155-176.

LASTRA, J.L.M., 2006. Semantic web services in factory automation: fundamental insights and research roadmap, Industrial Informatics, IEEE Transactions on, 2(1), pp 1-11.

LEE, C., KAO, Y., KUO, Y. and WANG, M., 2007. Automated ontology construction for unstructured text documents. Data & Knowledge Engineering, 60(3), 547-566.

LI YU, HUANG BIQING, LIU WENHUANG, GON HONGMEI and WU CHENG, 2000. Enterprise competence modelling and management, Systems, Man, and Cybernetics, 2000 IEEE International Conference on, vol 3, pp 2121-2126.

LI, ZHANJUN; RASKIN, VICTOR; RAMANI, KARTHIK, 2008, Developing Engineering Ontology for Information Retrieval, ASME Transactions Journal of Computing and Information Science in Engineering, 8 (1), 011003 (13 pages)

LIN, H.K. and HARDING, J.A., 2007. A manufacturing system engineering ontology model on the semantic web for inter-enterprise collaboration. Computers in Industry, 58(5), 428-437.

LINTHICUM, D.S., 1999. Enterprise application integration. Addison-Wesley Professional.

LUCZAK, H., 2005. Knowledge management in virtual organizations. Proceedings of ICSSSM '05. 2005 International Conference on Services Systems and Services Management, 2, pp 808-902.

MADNI, A.M., LIN, W. and MADNI, C.C., 2001. IDEONTM: An extensible ontology for designing, integrating, and managing collaborative distributed enterprises. Systems Engineering, 4(1), 35-48.

MARTIN, P. and EKLUND, P., 1999. Embedding knowledge in Web documents, May 11-14 1999, Elsevier North-Holland, Inc.

MENAGH, M., 1995. Virtues and vices of the virtual corporation. Computerworld, pp. 134, Available: http://careers.computerworld.com/111395.html

MERRILL, J., BAKKEN, S., ROCKOFF, M., GEBBIE, K. and CARLEY, K., 2007. Description of a method to support public health information management: organizational network analysis. Journal of Biomedal Informatics, 40(4), 422-428.

META DATA COALITION, 1999. Open Information Model, Version 1.1, Available: http://www.mdcinfo.com/OIM/index.html.

MILES, A., MATTHEWS, B., WILSON, M.D. and BRICKLEY, D., 2005. SKOS Core: Simple Knowledge Organisation for the Web, International Conference on Dublin Core and Metadata Applications, Madrid, Spain.

MISSIKOFF, M., SCHIAPPELLI, F. and TAGLINO, F., 2003. A Controlled Language for Semantic Annotation and Interoperability in e-Business Applications, 20-23 October 2003, pp1-6.

MORENO, J.L., 1937. Sociometry in relation to other social sciences. Sociometry, 1(2), 206-219.

MUATA, K., BRYSON, O. and GRANT, D. 2004, A manufacturing information systems architecture for the 21st century. ICEIS - INFORMATION SYSTEMS ANALYSIS AND SPECIFICATION, Porto, Portugal, 457-463.

MUTHAIYAH, S. and KERSCHBERG, L., 2008. Achieving interoperability in e-government services with two modes of semantic bridging: SRS and SWRL. Journal of Theoretical and Applied Electronic Commerce Research, 3(3), 52-63.

NEWMAN, M.E., 2001. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys Rev E Stat Nonlin Soft Matter Phys, 64(1-2),.

NOY, N.F. and MCGUNINESS, D., L., 2001. Ontology Development 101: A guide to creating your first ontology. Standford Knowledge Systems Laboratory Technical Report KSL-01-05 and Standford Medical Informatics Technical Report SMI-2001, Standford University.

NOY, N.F., SINTEK, M., DECKER, S., CRUBEZY, M., FERGERSON, R.W. and MUSEN, M.A., 2001. Creating Semantic Web contents with Protege-2000. Intelligent Systems, IEEE, 16(2), 60-71.

OPSAHL, T., AGNEESSENS, F. and SKVORETZ, J., 2010. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks, 32(3), 245-251.

PARETO, L. and SNIS, U.L., 2007. An Interactive Visualization Model for Competence Management: an Integrative Approach, 9th International Conference on Knowledge Management and Knowledge Technologies, Graz, Austria.

PATEL-SCHNEIDER, P.F. and HORROCKS, I., 2004. OWL: Web Ontology Language Semantics and Abstract Syntax. The World Wide Web Consortium, Available: http://www.w3.org/TR/owl-semantics/syntax.html.

PATIL, L., DUTTA, D. and SRIRAM, R., 2005. Ontology-based exchange of product data semantics. Automation Science and Engineering, IEEE Transactions on [see also Robotics and Automation, IEEE Transactions on], 2(3), 213-225.

PEDERSEN, J.D., 2007. A social network perspective on virtual organisations: social structure as enabler and barrier. International Journal of Networking and Virtual Organisations, 4(4), 431-445.

PÉPIOT, G., CHEIKHROUHOU, N., FURBRINGER, J. and GLARDON, R., 2007. UECML: Unified Enterprise Competence Modelling Language. Computers in Industry, 58(2), 130-142.

PLISSON, J., LJUBIC, P. and MOZETIC, I., 2007. An Ontology for Virtual Organization Breeding Environments. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 37(6), 1327-1341.

PRAHALAD, C.K. and HAMEL, G., 1990. The core competence of the corporation. Harvard Business Review, 68(3), 79-93.

RABARIJAONA, A., DIENG, R., CORBY, O. and OUADDARI, R., 2000. Building and searching an XML-based corporate memory. Intelligent Systems and their Applications, IEEE, 15(3), 56-63.

RAJSIRI, V., LORRÉ, J., BÉNABEN, F. and PINGAUD, H., 2007. Cartography for designing collaborative processes. In: R.J. GONÇALVES, J.P. MÜLLER, K. MERTINS and M. ZELM, eds, Enterprise Interoperability II. London: Springer, pp. 257-260.

RATCHEV, S., SHIAU, J. and VALTCHANOV, G., 2000. Distributed product and facility prototyping in extended manufacturing enterprises. Internation Journal of Production research, 38, 4495-4506.

RATCHEV, S., SHIAU, J. and VALTCHANOV, G., 2000. Distributed product and facility prototyping in extended manufacturing enterprises. Internation Journal of Production research, 38, 4495-4506.

RIOS-ALVARADO ANA, B., RAMÍREZ, R., CAROLINA, M. and MARCELÍN-JIMÉNEZ, R., 2009. A Semantic Web Approach to Represent and Retrieve Information in a Corporate Memory, OWL: Experiences and Directions, International Workshop, 23-24 October 2009|, Chantilly, Virginia, USA.

ROBINSON, J.P., 2004. What is the new economy. Alabama Cooperative Extension System, 1(4), 1-4.

SANCHEZ, R., 2004. Understanding competence-based management: Identifying and managing five modes of competence. Journal of Business Research, 57(5), 518-532.

SCHLENOFF, C., DENNO, P., IVESTER, R., LIBES, D. and SZYKMAN, S., 2000. An analysis and approach to using existing ontological systems for applications in manufacturing. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 14(4), 257-270.

SCHMIDT, A. and KUNZMANN, C., 2006. Towards a Human Resource Development Ontology for Combining Competence Management and Technology-Enhanced Workplace Learning. On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. pp. 1078-1087.

SCHOEMAKER, P., 1992. How to link strategic vision to core capabilities. Sloan Management Review, 534(1), 67-82.

SELL2WALES, CPV Codes-Guidance & Complete List. Available: https://www.sell2wales.co.uk/help/2120.html.

SHAPIRO, C. and VARIAN, H.R., 1998. Information Rules: A Strategic Guide to the Network Economy. Boston, MA: Harvard Business School Press.

SHEN, Z., MA, K.L. and ELIASSI-RAD, T., 2006. Visual Analysis of Large Heterogeneous Social Networks by Semantic and Structural Abstraction. IEEE Transactions on Visualization and Computer Graphics, 12(6), 1427-1439.

SIMAP, Tenders. Available: http://publications.europa.eu/tenders/index_en.htm.

SKYRME, D.J., 1999. Knowledge Networking: Creating the Collaborative Enterprise. Oxford, UK: Butterworth-Heinemann.

STEPHENSON, K., 2006. The Quantum Theory of Trust: Power, networks and the secret life of organisations. New Jersey: Pearson Education, FT Press.

STEVENS, R., WROE, C., BECHHOFER, S., RECTOR, A. and GOBLE, C., 2003. Building Ontologies in DAML+OIL. Comparative and Functional Genomics, 4(1), 133-141.

SUGUMARAN, V. and STOREY, V.C., 2002. Ontologies for conceptual modelling: their creation, use, and management. Data & Knowledge Engineering, 42(3), 251-271.

SUN MICROSYSTEMS, 2004. Identity management: Technology cornerstone of the virtual enterprise.

SWARNKAR, R., CHAUBE, A., HARDING, J.A., DAS, B.P. and TIWARI, M.K., 2009. Prioritising Tendering Activities for small to medium–sized enterprises (SMEs), June 2009, 15th International Conference on Concurrent Enterprising, Leiden, The Netherlands.

TAGG, R., 2001. Workflow in different styles of Virtual Enterprise, Information Technology for Virtual Enterprises, 2001. ITVE 2001. Proceedings. Workshop on, 2001, pp21-28.

TALLURI, S. and BAKER, R.C., 1996. Quantitative framework for designing efficient business process alliances, 1996, pp656-661.

TARASSOV, V., SANDKUHL, K. and HENOCH, B., 2006. Using ontologies for representation of individual and enterprise competence models, International Conference on Research, Innovation and Vision for the Future.

TERZIYAN, V. and KONONENKO, O., 2003. Semantic Web Enabled Web Services: State-of-Art and Industrial Challenges. ICWS-Europe 2003, LNCS 2853, 183-197.

TIDWELL, D., June, 2000. Web services-the web's next revolution, Available: http://ibm.com/developerworks/.

TOURTIER, P.A., 1995. Analyse préliminaire des métiers et de leurs interactions. GENIE, INRIA-Dassault-Aviation.

TUSHMAN, M.L. and O'REILLY, C.A., 2006. Ambidextrous organizations: Managing evolutionary and revolutionary change. In: D. MAYLE, ed, Managing innovation and change. 3rd edn. London, UK: Sage Publication Ltd, pp. 170-184.

USCHOLD, M. and JASPER, R., 1999. A Framework for understanding and classifying ontology applications, 1999, In Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods(KRR5), Stockholm, Sweden.

VAN HEIJST, G., VAN DER SPEK, R. and KRUIZINGA, E., 1996. Organizing Corporate Memories, 1996, B. Gaines and M. Musen eds pp1-17.

VARGAS-VERA, M., MOTTA, E., DOMINGUE, J., LANZONI, M., STUTT, A. and CIRAVEGNA, F., 2002. MnM: Ontology Driven Semi-automatic and Automatic Support for Semantic Markup. In: A. GÓMEZ-PÉREZ and V. BENJAMINS, eds, Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. Lecture Notes in Computer Science edn. Berlin/Heidelberg: Springer, pp. 213-221.

VARGAS-VERA, M., MOTTA, E., DOMINGUE, J., SHUM, S.B. and LANZONI, M., 2001. Knowledge extraction by using an ontology-based annotation tool, 2001, ACM pp5-12.

VERMA, A. and TIWARI, M.K., 2009. Role of corporate memory in the global supply chain environment. International Journal of Production Research, 47(19), 5311-5342.

VERNADAT, F., 2002. UEML: towards a unified enterprise modelling language. International Journal of Production Research, 40(17), 4309-4321.

VERNADAT, F.B., 2002. Enterprise modelling and integration (EMI): Current status and research perspectives. Annual reviews in control, 26(1), 15-25.

WANG, D., YUNG, K. and IP, W., 2002. Partner selection model and soft computing approach for dynamic alliance of enterprises. Science in China Series F: Information Sciences, 45(1), 68-80.

WANG, T., CHEN, L.Y. and CHEN, Y., 2007. Applying Incomplete Linguistic Preference Relations to Tender Selection, International Conference on Wireless Communications, Networking and Mobile Computing, WiCom 2007, 2007, pp6734-6737.

WANG, X. and XIONG, G., 2001. Design rationale as part of corporate technical memory. Systems, Man, and Cybernetics, 2001 IEEE International Conference on, 3, 1904-1908 vol.3.

WATTS, D.J., 2003. Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton, New Jersey: Princeton University Press.

WEBER, C.A., CURRENT, J.R. and BENTON, W.C., 1991. Vendor selection criteria and methods. European Journal of Operational Research, 50(1), 2-18.

WICK, C.W., 2001. Teaching an old economy company new economy tricks: knowledge management at a multinational information technology service firm. PhD Thesis edn. Texas: Texas Tech University.

WONG, C.H., HOLT, G.D. and HARRIS, P., 2001. Multi-criteria selection or lowest price? Investigation of UK construction clients' tender evaluation preferences. Engineering, Construction and Architectural Management, 8(4), 257-271.

XING JIANG and AH-HWEE TAN, 2005. Mining ontological knowledge from domain-specific text documents.

XU, H. and XIAO, D., 2007. Applying Semantic Web Services to Automate Network Management. Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on, 461-466.

YALAN YAN, JINLONG ZHANG and MI YAN, 2006. Ontology Modelling for Contract: Using OWL to Express Semantic Relations.

YANG, H., 2009. Automatic generation of semantically enriched web pages by a text mining approach. Expert Systems with Applications, 36(6), 9709-9718.

YIT CORPORATION, F., Jan 2000. GLOBEMEN- Global Engineering and Manufacturing in Enterprise Networks, Available: http://globemen.vtt.fi/.

YU, B., HARDING, J.A. and POPPLEWELL, K., 2000. Supporting enterprise design through multiple views. International Journal of Agile Management Systems, 2, 71-83.

ZHANG, J., 1998. An Architectural Approach to Modelling the Extended Enterprise. PhD Thesis edn. Galway, Ireland: National University of Ireland.

ZHOU, J. and DIENG-KUNTZ, R., 2004. Manufacturing ontology analysis and design: towards excellent manufacturing. 2nd IEEE International Conference on Industrial Informatics, INDIN '04., 39-45.

ZIMMERMANN, F.O., 1997. Structural and Managerial Aspects of Virtual Enterprises, April 7-10, Proceedings of the European Conference on Virtual Enterprises and Networked Solutions – New Perspective on Management, Communication and Information Technology, April 7-10, Paderborn, Germany.

APPENDIX A

List of Publications

The following papers have been published or accepted for publication related to this research.

1. **Semantic Web in Manufacturing**

   Nitesh Khilwani, J. A. Harding, A. K. Choudhary

   (Published in IMechE: Part B: Journal of Engineering Manufacture, 2009)

   Link- http://journals.pepublishing.com/content/j2767r260p731527/

   Abstract- Advances in manufacturing systems include attempts to create collaborative networks for enterprise integration and information interoperability. To achieve collaboration and sharing effectively, various networking technologies have been proposed in the literature. The web has emerged as a basic entity for interconnecting man and machine and almost all parts of the enterprise community are being reshaped to exploit the opportunities that it offers. Apart from web technology, there are various other tools and techniques that have attracted research communities for representing data in ways that both machines and humans can understand. Semantic web, the second-generation web technology, is enriched by machine-processable information to support the users in their tasks. This paper presents the vision of the semantic web and describes ontologies and associated metadata as the building blocks of the semantic web. It reviews the literature dealing with the application of the semantic web and ontology in the broad domain of manufacturing. First, brief details about key enablers, i.e. web services, semantic web, semantic services, and ontology, are presented. Then the implementation of these approaches in different sectors of manufacturing is discussed. A knowledge base for all the information resources concerned with the manufacturing domain is also built up in this paper. An ontology model for a knowledge base of information resources is designed in Protege software, which can be used for storing and searching information about authors, journals, blogs, newspapers, and many other sources of information.

2. **ECOS: Publishing the published competences**

Nitesh Khilwani, J. A. Harding, M. K. Tiwari

Abstract- Competence is a standardized way to define profile of an enterprise. Understanding and auditing competencies acquired, required and desired by a company and further representing them in a structured manner is a beneficial step for enhancing their performance. Ontology is emerging as an effective tool to structure competences for comprehensive and transportable machine understanding. In this paper, ECOS (Enterprise Competence Organization Schema) is presented as a mechanism to capture enterprise competence in a computer understandable manner. The objective behind this concept is to create a web of machine readable pages describing basic information and competences of enterprises with sets of interconnected data and semantic models. The ECOS ontology captures enterprise competences using a consistent and comprehensive list of concepts and vocabulary and converts them into a semantic web resource using the Web Ontology Language (OWL). The novel concept of an ECOS card and ECOS form is proposed and used for developing and publishing enterprise competences. Examples from real life enterprise applications of ECOS are also shown in this paper.

3. **Role of Semantic Web in the changing context of Enterprise Collaboration**

Nitesh Khilwani, J. A. Harding

Abstract- In order to compete with the global giants, enterprises are concentrating on their core competencies and collaborating with organizations that compliment their skills and core activities. The current trend is to develop temporary alliances of independent enterprises, in which companies can come together to share skills, core competencies and resources. However, knowledge sharing and communication among multidiscipline companies is a complex and challenging problem. In a collaborative environment, the meaning of knowledge is drastically affected by the context in which it is viewed and interpreted; thus necessitating the treatment of structure as well as semantics of the data stored in enterprise repositories. Keeping the present market and technological scenario in mind, this research aims to propose tools and techniques that can enable companies to assimilate distributed information resources and achieve their business goals.

4. **Ontology mining for platform extraction in product development**

Nitesh Khilwani, J. A. Harding, M. K. Tiwari

(Published in Proceedings of the 14th International Conference on Concurrent Enterprising, ICE 2008)

Link- http://www.ice-proceedings.org/apps/pub.asp?Q=3312

Abstract- In this paper, a Product-Process Ontology is developed to capture, design and organize the product data knowledge in a manner that computer can use and manipulate it meaningfully. It is proposed to enhance the manageability of large amount of product data and ultimately influence the basic organization's functions such as customer order processing, variety management, production planning, material and capacity planning etc. Moreover, an ontology mining technique is proposed to exploit the available product data information for designing a platform that offers minimum production cost, without hampering the performance of the product. As a demonstration, five products from a family of circular saw is considered to validate the applicability and robustness of proposed architecture.

APPENDIX B

Standard Classifications

Classifications gather and organize information meaningfully and systematically into a standard format that can be useful for determining the similarity of ideas, events, objects or persons. The classification means the creation of an exhaustive and structured set of mutually exclusive and well-described categories, generally presented as a hierarchy of numeric or alphabetical codes assigned to them. Different international bodies have given classifications on matters such as economics, demographics, labour, health, education, social welfare, geography, environment and tourism. Some of the standard classifications used in this paper are:

- ISIC http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=17 (International Standard for Industrial Classification): It is a standard developed by the United Nations Statistical Division for classifying productive economic activities (i.e. the same raw materials, process of production, skills or technology) . The ISIC system is now used widely by governments and international bodies as a way if classifying data according to economic activity. One key purpose of the code is to standardize data collection and promote international comparability. The entire ISIC classification is based on four levels, where level 1 includes 17 sections identified by alphabetical letters A to Q, level 2 covers 62 divisions identified by two-digit numerical codes, level 3 is defined by 161 groups identified by three-digit numerical codes and level 4 includes 298 classes represented with 4 digit code. For example:

    Sector: I - Transport, storage and communications

    Division: 64 - Post and telecommunications

    Group: 641 - Post and courier activities

    CLASS: 6411 - NATIONAL POST ACTIVITIES

- CPC   http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=16   (Central   Product Classification), defined by UN Statistical Division, constitutes a comprehensive classification of all goods and services. CPC presents categories for all products that can be the object of domestic or international transactions or that can be entered into stocks. It includes products that are an output of economic activity, including transportable goods, non-transportable goods and services. The classification structure of CPC comprises: Sections - one digit code; Divisions - two-digit code; Groups - three-digit code; Classes - four-digit code; Subclasses – five-digit code. For example,

    Sector: 8 - Business and production services

    Division: 84 - Telecommunications, broadcasting and information supply services

    GROUP: 841 - TELEPHONY AND OTHER TELECOMMUNICATIONS SERVICES

    CLASS: 8411 - CARRIER SERVICES

    SUBCLASS: 84110 - CARRIER SERVICES

- NUTS   http://ec.europa.eu/eurostat/ramon/nuts/basicnuts_regions_en.html (Nomenclature of Territorial Units for Statistics), developed by the European Union, is a geocode standard for referencing the administrative divisions of countries. This code begins with a two-letter code referencing the country, followed by 3 level codes. The three levels represent NUTS 1 (e.g. region, states etc), NUT 2 (e.g. county, province, etc) and NUT 3 (city, area, etc) respectively.

    UK: UNITED KINGDOM

    UKF: EAST MIDLANDS, ENGLAND

    UKF2: LEICESTERSHIRE, RUTLAND AND NORTHAMPTONSHIRE

    UKF22: LEICESTERSHIRE CC AND RUTLAND

- ISCO   http://www.ilo.org/public/english/bureau/stat/isco/index.htm   (International Standard Classification of Occupations) is standard structure developed by UN Statistical Division. This classification organizes jobs into a clearly defined set of

groups according to the tasks and duties undertaken in the job. The classification structure of ISCO includes 4 levels: Level 1- Major groups (10), Level 2: Sub-major groups (28), Level 3- Minor groups (118) and Level 4- Unit groups (390).

MAJOR: 1 – MANAGERS

      SUB-MAJOR: 12 - ADMINISTRATIVE AND COMMERCIAL MANAGERS

      MINOR: 121 - BUSINESS SERVICES AND ADMINISTRATION MANAGERS

      UNIT: 1211 - FINANCE MANAGERS

- CPV http://www.cpvclassification.com/ (Common Procurement Vocabulary) was created in 1996 as a tool for improving transparency and efficiency in the field of public procurement. Use of standard terms in the CPV makes it easier for buyers and sellers to identify the tenders in which they are interested. This vocabulary is based on a tree structure comprising codes of up to nine digits associated with a wording that describes the supplies, works or services forming the subject of the contract.
    - The first two digits identify the divisions (XX000000-Y)
    - The first three digits identify the groups (XXX00000-Y)
    - The first four digits identify the classes (XXXX0000-Y)
    - The first five digits identify the categories (XXXXX000-Y)

The former classifications are used in ECOS model for providing a standard vocabulary for competence representation.

APPENDIX C

VCARD Ontology

vCard (aka. Versit card) is a standard file format for electronic business cards. It was originally proposed in 1995 by the Versit consortium including Apple, AT&T Technologies, IBM and Siemens. vCard is powerful means of Personal Data Interchange (PDI) used for exchanging information on emails, web etc. This is an electronic form of traditional business cards containing basic information such as name, address, phone numbers, URLs, logos, photographs, audio clips etc. In 2001, IPR Systems Pty Ltd specified a RDF (Resource Description Framework) framework for vCard profile. Resource Description Framework (RDF) is a W3C standard used for defining meta-data model. Iannella, R., used this language for developing a metadata model of vCard for semantic web [63]. In this paper, RDF framework of vCard profile is used to represent ecos:Address in the ECOS ontology model. An example of vCard schema used in ECOS model is as follows:

```
<ecos:Address>
    <vCard:ADR rdf:parseType="Resource">
        <vCard:Street>Asbhy Rpad </vCard:Street>
        <vCard:Locality>Loughborough</vCard:Locality>
        <vCard:Pcode>LE11 3Tu</vCard:Pcode>
        <vCard:Country>United Kingdom</vCard:Country>
    </vCard:ADR>
</ecos:Address>
```

APPENDIX D

An RDF file generated for the text mentioned in Figure 6-7.

```xml
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:kmm="http://www.name.com/">
<rdf:Description rdf:about="http://www.name.com/doc#number">
  <kmm:section rdf:parseType="Resource">
    <kmm:sen rdf:parseType="Resource">
      <kmm:id>#0</kmm:id>
      <kmm:svo rdf:parseType="Resource">
        <kmm:s>Internation Business Machines Corporatation (IBM)</kmm:s>
        <kmm:v>develops</kmm:v>
        <kmm:v>manufactures</kmm:v>
        <kmm:o>information technology products and services worldwide.</kmm:o>
      </kmm:svo>
      <kmm:phrase>develops   and   manufactures   information   technology   products   and
services</kmm:phrase>
      <kmm:keyword>IBM</kmm:keyword>
      <kmm:keyword>International Business Machines Corporation</kmm:keyword>
      <kmm:term>information technology products and services</kmm:term>
      <kmm:term>worldwide</kmm:term>
      <kmm:verb>develops</kmm:verb>
      <kmm:verb>manufactures</kmm:verb>
      <kmm:text>International   Business   Machines   Corporation   (IBM)   develops   and
manufactures information technology products and services worldwide.</kmm:text>
    </kmm:sen>

    <kmm:sen rdf:parseType="Resource">
      <kmm:id>#1</kmm:id>
      <kmm:svo rdf:parseType="Resource">
        <kmm:s>IBM</kmm:s>
        <kmm:v>was founded</kmm:v>
        <kmm:o>in 1910.</kmm:o>
      </kmm:svo>
      <kmm:svo rdf:parseType="Resource">
        <kmm:s>IBM</kmm:s>
        <kmm:v>is based in</kmm:v>
        <kmm:o>Armonk, New York</kmm:o>
      </kmm:svo>
      <kmm:iphrase>founded in 1910</kmm:iphrase>
      <kmm:iphrase>based in Armonk, New York</kmm:iphrase>
      <kmm:number>1910</kmm:number>
      <kmm:keyword>IBM</kmm:keyword>
      <kmm:keyword>Armonk</kmm:keyword>
      <kmm:keyword>New York</kmm:keyword>
      <kmm:verb>founded</kmm:verb>
      <kmm:verb>based</kmm:verb>
      <kmm:text>IBM was founded in 1910 and is based in Armonk, New York.</kmm:text>
    </kmm:sen>
  </kmm:section>
</rdf:Description></rdf:RDF>
```

APPENDIX E

List of Software used in this research (in alphabetical order).

1   Apache Lucene   http://lucene.apache.org/
     Apache Lucene is a high-performance, full-featured text search engine library written entirely in
     Java. It is a technology suitable for nearly any application that requires full-text search, especially
     cross-platform.

2   GWT   http://code.google.com/webtoolkit/
     Google Web Toolkit is an open source set of tools that allows web developers to create and
     maintain complex JavaScript front-end applications in Java.

3   JAMA   http://math.nist.gov/javanumerics/jama/
     JAMA, A Java Matrix Package, is a basic linear algebra package for Java. It provides user-level classes
     for constructing and manipulating real, dense matrices.
     It is meant to provide sufficient functionality for routine problems, packaged in a way that is natural
     and understandable to non-experts.

4   Jena API   http://jena.sourceforge.net/
     It is a Java framework for building Semantic Web applications. It provides a programmatic
     environment for RDF, RDFS and OWL, SPARQL and includes a rule-based inference engine.
     Jena is open source and grown out of work with the HP Labs Semantic Web Programme.

5   Jena TDB   http://openjena.org/TDB/
     TDB is a component of Jena. It provides for large scale storage and query of RDF datasets using a
     pure Java engine. TDB supports SPARQL.

6   Stanford NLP Parser   http://nlp.stanford.edu/software/lex-parser.shtml
     A natural language parser is a program that works out the grammatical structure of sentences, for
     instance, which groups of words go together (as "phrases") and which words are the subject or
     object of a verb.
     This package is a Java implementation of probabilistic natural language parsers, both highly
     optimized PCFG and lexicalized dependency parsers, and a lexicalized PCFG parser.

7   Treebolic   http://treebolic.sourceforge.net/
     Treebolic is a Java component (widget) whose purpose is to provide a hyperbolic rendering of
     hierarchical data.

APPENDIX F

The results evaluated for the three centrality measures (degree, closeness and betweenness) for ECOS-web graph in Figure 8-16.

Enterprise	Degree	Enterprise	Betweenness	Enterprise	Closeness
13	3.441	17	49	5	3.038
6	3.142	3	45	19	3.036
10	3.053	4	38	13	2.983
5	3.022	13	33	6	2.922
20	2.864	18	31	16	2.907
3	2.783	21	31	4	2.847
17	2.76	24	31	10	2.832
21	2.582	6	30	3	2.817
4	2.409	5	25	23	2.765
25	2.36	8	22	20	2.729
11	2.234	25	21	12	2.7
12	2.232	11	20	14	2.626
9	2.203	20	11	22	2.608
23	2.157	10	10	9	2.559
24	2.142	7	8	21	2.524
14	2.131	23	7	8	2.495
1	2.094	16	6	24	2.475
19	2.05	12	5	18	2.444
8	1.784	1	4	25	2.324
2	1.686	9	4	17	2.309
15	1.521	14	4	1	2.25
18	1.131	19	3	11	2.227
16	1.07	2	0	15	2.221
7	0.69	15	0	2	2.044
22	0	22	0	7	2.042